



Transcriptional Regulation of Nodal Target Genes in Early Zebrafish Development

Citation

Akhmetova, Laila. 2016. Transcriptional Regulation of Nodal Target Genes in Early Zebrafish Development. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:33493493>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Transcriptional regulation of Nodal target genes in early
zebrafish development

A dissertation presented

by

Laila Akhmetova

to

The Department of Molecular and Cellular Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biochemistry

Harvard University

Cambridge, Massachusetts

April 2016

© 2016 Laila Akhmetova

All rights reserved.

Transcriptional regulation of Nodal target genes in early zebrafish development

ABSTRACT

Nodal signaling is one of the principal players in the process of gastrulation, during which the primary germ layers (endoderm, mesoderm, and ectoderm) are formed and organized in their proper locations. During my PhD I studied how Smad2 and FoxH1 transcription factors regulate the expression of Nodal target genes, and the relationship between the chromatin state of target genes and their expression.

In order to carry out in depth analysis of FoxH1 function I generated a complete mutant of this transcription factor and used deep sequencing to identify which genes FoxH1 regulates in zebrafish development. Using ChIP-Seq experiments, I also found the binding sites of FoxH1 and found that FoxH1 is capable of binding to genomic DNA in the absence of Nodal signaling. This finding suggests that it may act as a pioneer factor, preparing target genes for rapid activation when gastrulation starts. I also identified 54 direct FoxH1 target genes that are not Nodal-dependent, as well as 13 genes that are repressed, rather than activated, by FoxH1.

To identify Smad2 binding sites, I carried out ChIP-Seq in embryos overexpressing Nodal signal Squint, thus detecting loci that bind Smad2 after exposure to high Nodal levels. I tested the identified Smad2-bound DNA elements for their gene regulatory potential, and discovered that they are sufficient to drive gene expression in a Nodal-dependent manner. I also identified 26 previously unpublished

Nodal target genes, and 55 genes that are bound by Smad2 upon exposure to high, but not low levels of Nodal signaling.

In the last chapter I describe the study of the interaction between Nodal signaling in early zebrafish development and chromatin marks. I found that exposure of embryonic cells to high levels of Nodal is associated with low levels of H3K27me3 and high levels of H3K4me3 marks on Nodal target genes, compared to unexposed cells. I also describe a Cas9-based system that we used to change H3K27me3 levels in a targeted manner, and tested it in a developing embryo on a Nodal-responsive *fgf8a* gene. Our results suggest that reduction of H3K27me3 mark on its own is not sufficient to affect the expression of this gene, and additional mechanisms are involved in target gene activation by Smad2.

Table of Contents

Acknowledgements	VI
List of figures	VII
List of tables	IX
Chapter I	1
Introduction	
Chapter II	26
Analysis of FoxH1 function and target genes	
Chapter III	51
Identification of Smad2 binding sites and regulatory elements	
Chapter IV	74
Analysis of the relationship between Nodal target gene expression and chromatin modifications	
Chapter V	96
Perspectives and outstanding questions	

Acknowledgments

People I want to thank for enriching my experience in graduate school and making the years spent in Cambridge the most interesting and stimulating in my life so far (incomplete list)

- Alex Schier: for being my mentor, for inspiring me as a scientist, as a PI, and as a person. For building a lab where one can't help but learn and be excited about research.
- The members of my committee. Current: Susan Mango, Sharad Ramanathan, John Rinn and past: Andrew McMahon, Erin O'Shea, Nicole Francis. For letting me learn from you and supporting me through the years.
- People in the Schier lab, past and present: Steve Zimmerman, Jamie Gagnon, Meg Norris, Shristi Pandey, Michal Rabani, Owen Randlett, Constance Richter, Eivind Valen, Ian Woods, David Schoppik, Tessa Montague, Jeff Farrell, Guo-Liang Chew, Mehdi Goudarzi, Martin Haesemeyer, Peng Huang, Farhad Imam, Alix Lacoste, Jen Li, Florian Merkle, Albert Pan, Jason Rihel, Drew Robson, Patrick Muller, Nathan Lord. For being some of the most interesting, intelligent, fun, helpful, knowledgeable, kind people in the world.
- Special thanks go to: Summer Thyme (for being a great friend and collaborator and for always laughing at my jokes); Julien Dubrulle (for teaching me so much about experimental biology and French soccer teams); Nadine Vastenhouw (for guidance and inspiration in my ChIP endeavors); Andrea Pauli (for always being ready to help, even at 3am when she is working towards her own deadline next morning). Last but not the least, to Katherine Rogers (for being one of my best friends, for watching Game of Thrones with me, and commiserating when experiments didn't work).
- Xiuye Chen and Hannah Shen, for being there for me through all the ups and downs and in-betweens of graduate school life.
- My Mom, my Dad and my sister: for your unconditional love.
- My cat Yuki. For being my cat.

List of figures

CHAPTER I

Figure 1.1	3
<i>Schematic representation of a 5-hour post fertilization (5 hpf) zebrafish embryo</i>	
Figure 1.2	5
<i>Wolpert's model of morphogen systems</i>	
Figure 1.3	6
<i>Intracellular transduction of Nodal signaling</i>	
Figure 1.4	7
<i>Smad2 is necessary for mesendoderm specification by Nodal signaling</i>	
Figure 1.5	9
<i>FoxH1 expression in early zebrafish embryo</i>	
Figure 1.6	11
<i>Characteristics of a pioneer factor</i>	
Figure 1.7	12
<i>Forkhead factors possess structural features important for their pioneer factor role</i>	
Figure 1.8	14
<i>DNA-binding domain of FoxH1 is able to bind to DNA independent of Nodal signaling</i>	
Figure 1.9	16
<i>Chromatin structure and its modifications reflect and govern the transcriptional state of the associated genes</i>	

CHAPTER II

Figure 2.1	28
<i>FoxH1 transcription factor in zebrafish and its published mutants</i>	
Figure 2.2	29
<i>FoxH1 morphant exhibits an early gastrulation phenotype</i>	
Figure 2.3	30
<i>FoxH1 mutant generation using Cas9-mediated mutagenesis</i>	
Figure 2.4	31
<i>Zygotic homozygous FoxH1 E1 mutant embryo</i>	

Figure 2.5	32
<i>FoxH1 maternal-zygotic E1 mutant phenotype</i>	
Figure 2.6	33
<i>Side-by-side comparison of the MZ midway mutant and MZE1 mutant</i>	
Figure 2.7	30
<i>FoxH1 mutant generation using Cas9-mediated mutagenesis</i>	
Figure 2.8	35-36
<i>ChIP tracks for FoxH1 transcription factor binding and enhancer chromatin marks</i>	
Figure 2.9	37
<i>FoxH1 peaks de-novo motif analysis</i>	

CHAPTER III

Figure 3.1	55-56
<i>Smad2 transcription factor binds to regulatory elements in a Nodal-dependent manner</i>	
Figure 3.2	58
<i>Binding motifs enriched in genomic sequences bound by Smad2 in Squint-injected embryos</i>	
Figure 3.3	60
<i>Smad2-bound putative enhancer elements drive gene expression in the embryo margin zone</i>	
Figure 3.4	62
<i>Expression driven by the regulatory element of ntla gene is responsive to Nodal signaling levels</i>	
Figure 3.5	64
<i>Overlap between Smad2-proximal genes identified based on the Wardle group data, my ChIP-Seq data, and Nodal-responsive genes</i>	

CHAPTER IV

Figure 4.1	78-79
<i>Genome-wide chromatin modification changes in zebrafish embryo in response to Nodal signaling</i>	
Figure 4.2	80
<i>Induced gene expression using targeted dCas9-VP64 localization</i>	
Figure 4.3	81
<i>Changes in chromatin marks induced by targeted dCas9-VP64 localization</i>	

Figure 4.4	83
<i>H3K7me3 mark reduction by dCas9-UTX</i>	

Figure 4.5	84
<i>Combinations of 4 gRNAs are sufficient for H3K7me3 mark reduction</i>	

Figure 4.6	85
<i>fgf8a gene expression levels following the injection of dCas9-sfGFP or dCas9-UTX construct with 4 gRNAs</i>	

List of tables

Table 2.1	39
<i>FoxH1 direct target genes, based on RNA-Seq and ChIP-Seq analysis</i>	

Table 3.1	64
<i>Previously unpublished Nodal target genes</i>	

Dedication
To my family

CHAPTER I

INTRODUCTION

*It is not birth, marriage, or death, but gastrulation,
which is truly the most important time in your life.*

Lewis Wolpert (1986)

Embryonic development starts with the fertilization of the egg, formation of a pluripotent cell population, and their subsequent differentiation into distinct lineages. This complex process is regulated by the interaction of multiple signaling pathways and transcriptional regulators. Nodal signaling is one of the principal players in the process of gastrulation, during which the primary germ layers (endoderm, mesoderm, and ectoderm) are formed and organized in their proper locations.

In this introduction I will discuss the roles of Nodal signaling during early vertebrate development and the factors that are involved in the transduction and regulation of this pathway. I will talk about the nature of morphogens, and how it relates to Nodal signaling. Special attention will be paid to two important transcription factors in this pathway: Smad2 and FoxH1. I will discuss how these factors regulate the expression of Nodal target genes, and the relationship between the chromatin state of target genes and their expression. I will pose some of the unresolved questions in the field of transcriptional regulation of Nodal-responsive genes, and explain how I addressed them in my research.

Nodal is a conserved early developmental signal necessary for mesendoderm induction and left-right specification

Some of the most important signaling pathways in animal development are those belonging to the TGF-beta superfamily. It is a large family of polypeptides with similar structures and similar intracellular signal transduction machinery. This superfamily comprises of at least 42 members in vertebrates, and includes TGF-betas, bone morphogenic proteins (BMPs), growth and differentiation factor (GDF), myostatin, and Nodal signals (Oshimori and Fuchs, 2012). Nodal subfamily is considered to be one of the most evolutionary ancient of the TGF-beta superfamily members (Pang et al., 2011), and it is present in most metazoans, though notably absent from *Drosophila* and *C. elegans* (Rebagliati et al., 1998).

Nodal was one of the first genes knocked out in mice (Zhou et al., 1993). A retrovirally induced mutation of this gene in mouse led to a failure of mesoderm formation. Instead of giving rise to mesoderm, embryonic ectoderm in these mutants overproliferated and then rapidly degenerated. Expression of this gene in mouse embryo is highly localized in the node at the anterior of the primitive streak, which gave the Nodal gene its name (Zhou et al., 1993). The node in the mouse is analogous to chick Hensen's node, dorsal lip in *Xenopus*, and embryonic shield in zebrafish. Mice, as well as humans and birds, have one copy of the Nodal gene, while zebrafish have three: *squint*, *cyclops*, and *southpaw* (Erter et al., 1998; Feldman et al., 1998a; Rebagliati et al., 1998; Sampath et al., 1998). From its first discovery in 1993, Nodal signaling has been shown to play crucial role in mesendoderm induction, left-right patterning, and self-renewal in embryonic stem cells (Conlon et al., 1994; Schier and Shen, 2000; Schier, 2009). In the adult organisms Nodal plays an important role

in tissue homeostasis and Nodal signaling deregulation is implicated in tumorigenesis and metastasis (Strizzi et al., 2012).

Two of the three Nodal signals, Squint and Cyclops, are expressed during gastrulation and dictate mesendoderm formation (Erter et al., 1998; Feldman et al., 1998a). These signals are produced at the blastula margin, along with their inhibitor Lefty [Fig.1.1]. The third signal Southpaw is expressed later in the development and is necessary for the left-right specification. Nodal signals are produced as proproteins, and require the action of proteases to reach their mature active form (Le Good et al., 2005).

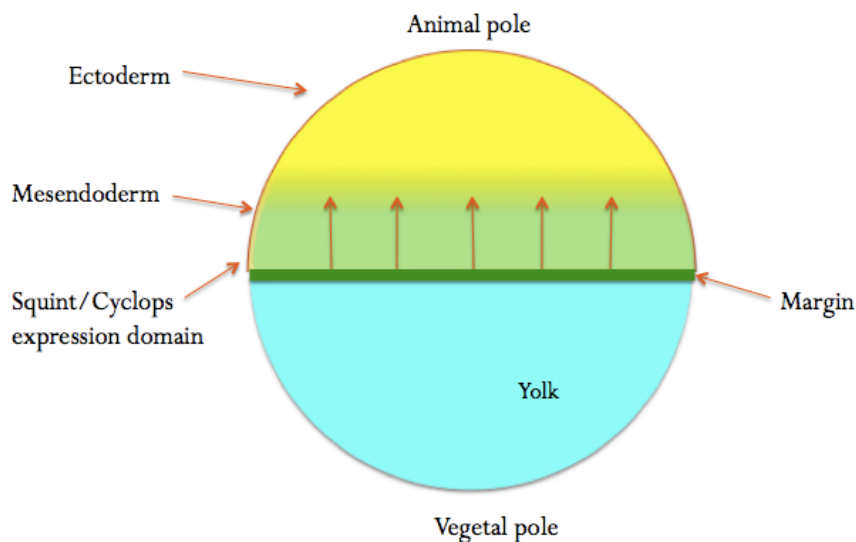


Figure 1.1

Schematic representation of a 5-hour post fertilization (5hpf) zebrafish embryo

Nodal signals (Squint and Cyclops) are produced at the margin and form an animal-vegetal concentration gradient, inducing mesendoderm formation.

Nodal acts as a morphogen signal

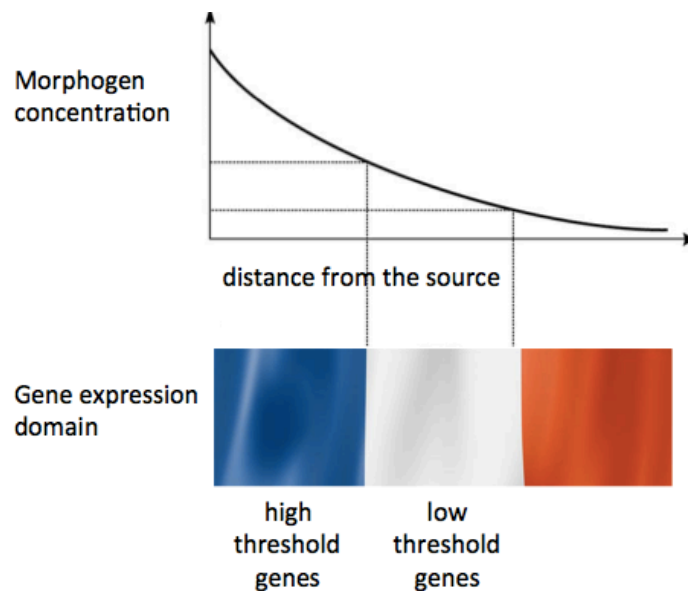
Unlike many signaling systems, which elicit a binary on/off response, Nodal signaling belongs to the class of morphogens. Morphogen signals diffuse in the extracellular space and can act directly on cells away from the source, producing different responses depending on the level of the exposure to the signal.

How morphogen gradients are formed and then translated into differential target gene expression remains an active research area. In his paper written in 1952, Alan Turing proposed a chemical mechanism for the formation of biological patterns (morphogenesis) (Turing, 1952), describing a reaction-diffusion system that can establish a morphogen gradient in an initially uniform signaling field.. In simplistic terms, Turing's original morphogens can be short-range activators coupled with long-range inhibitors, each of which acts on itself as well as the other (Gierer and Meinhardt, 1972). In the 1960s Lewis Wolpert proposed a positional information-based model of how a morphogen could subdivide a tissue into domains of different target gene expression (Wolpert, 1969). Such a system has preexisting asymmetry that gives rise to a morphogen gradient [Fig.1.2]. While these models were often considered to be mutually exclusive, they can in fact be collaborating to establish morphogen response systems (Green and Sharpe, 2015). In particular, the pattern formed by Nodal and Lefty is a single smooth gradient that is capable of providing positional information to every cell, and looks therefore like the Wolpert model. However, the gradient itself is created by Turing dynamics, where the short-range Squint and Cyclops signals interact with the long range Lefty inhibitors (Muller et al., 2012).

The established morphogen gradient must then be translated into differential gene expression. Classic models of morphogen interpretation, such as the one for Bicoid in *Drosophila*, are centered on regulation through differential DNA affinity (Driever and Nusslein-Volhard, 1988). In this model the regulatory elements of low threshold genes have higher affinity for activating factors, and vice versa. Recent studies, however, reveal that temporal dynamics play an important role in the regulation of Nodal signaling targets (Dubrulle et al., 2015; van Bortel et al., 2015).

These studies show that intrinsic transcription levels of a gene and delays in transcriptional onset can have a strong effect on the range of gene expression, and that temporal information can be converted into spatial information in the developing embryo.

A.



B.

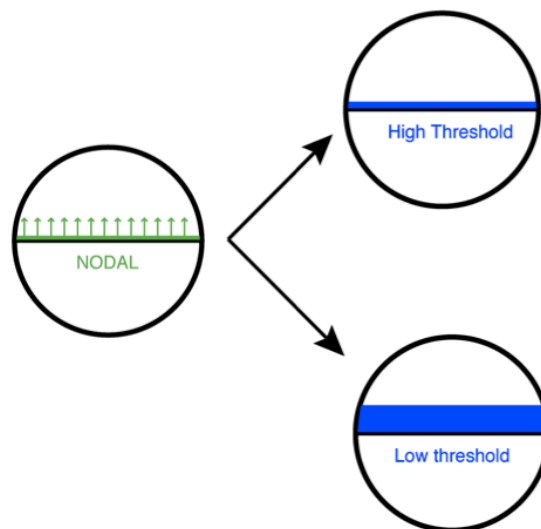


Figure 1.2

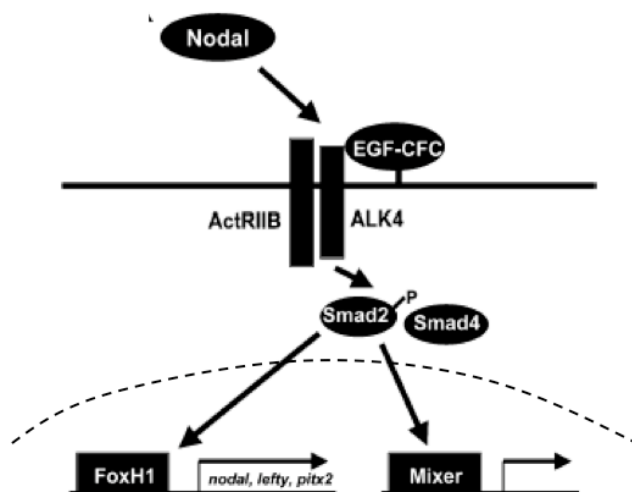
Wolpert's model of morphogen systems

A) In Wolpert's positional information model, also known as the French flag model, a prior asymmetry results in a graded monotonic distribution of a morphogen. Cells then use this distribution to make fate choices, where the concentration and/or the duration of the signal that a cell is exposed to dictates the expression of a specific gene/genes. Adapted from (Wolpert, 1969).

B) Nodal signaling in a zebrafish embryo forms a concentration gradient starting at the yolk margin. At least two groups of Nodal-responsive genes have been identified: high threshold (*gsc*, *sox32*) and low threshold (*ntla*, *flh*) (Schier and Shen, 2000).

Intracellular Nodal signaling transduction

As with the other TGF-beta superfamily signals, transduction of Nodal signaling happens through the phosphorylation of specific serine-threonine kinases. Nodal ligands interact with two types of transmembrane receptors, which have intrinsic kinase activities in their cytoplasmic domain (Wrana et al., 1994). The signaling molecules bind to type II Activin receptors (ActRII/IIB), which leads to the recruitment and activation of type I Activin receptors (ALKs) (Tsuchida et al., 2004) [Fig.1.3]. Another group required for Nodal signaling in some systems are extracellular EGF-CFC proteins, thought to act as coreceptors (Shen and Schier, 2000). In zebrafish, the loss of EGF-CFC protein one-eyed pinhead (*oep*) leads to complete loss of Nodal signaling and failure to develop most mesendodermal tissues (Gritsman et al., 1999). Successful activation of the Nodal receptors then leads to the phosphorylation of Smad2/3 transcription factors and their translocation into the nucleus [Fig.1.3] (Massague et al., 2005).



adapted from (Schier and Shen, 2000)

Figure 1.3

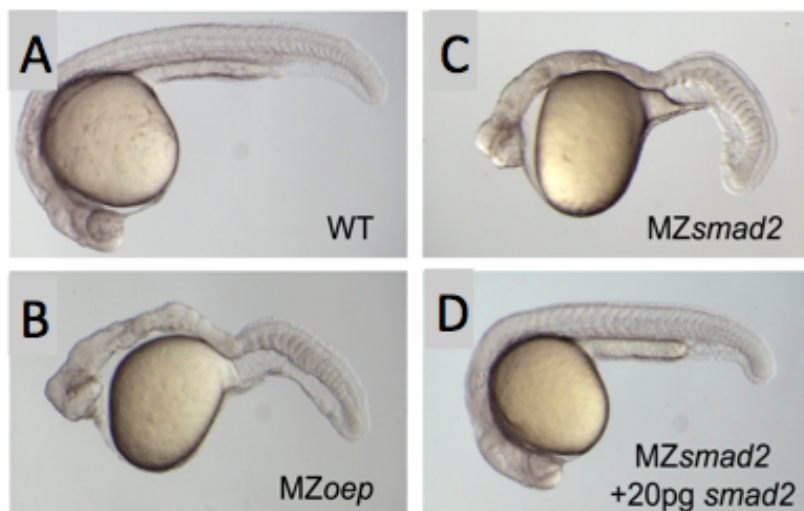
Intracellular transduction of Nodal signaling

Binding of the Nodal ligand to the ActRIIB/ALK4 receptors and EGF-CFC coreceptors leads to the activation of Smad transcription factors and their translocation into the nucleus, where they activate target genes.

Transcription factors in Nodal signaling pathway

The canonical pathway for all TGF-beta superfamily members involve the activation of receptor-regulated Smads (R-Smads), with Smads1/5/8 being involved in the transduction of the BMP and GDF signals and Smad2/3 – Nodal, Activin and TGF-beta signals (Massague, 2012). In their turn, inhibitory Smads 6 and 7 provide negative feedback (Massague et al., 2005).

While both Smad2 and Smad3 proteins are expressed in zebrafish embryo, early in gastrulation only Smad2 is present and is the main Nodal signal transducer necessary for mesendoderm development (Dubrulle et al., 2015). R-Smads can enter the nucleus either as homomeric complexes or in association with Smad4, although Smad4 mutant phenotypes are not as severe as those of Smad2/3 loss (Chu et al., 2005). Smad2 is provided maternally in zebrafish and its loss leads to a phenotype identical to that of complete Nodal signaling loss, where endoderm and head and trunk mesoderm are absent [Fig.1.4] (Dubrulle et al., 2015) (Feldman et al., 1998b).



from (Dubrulle et al., 2015)

Figure 1.4

Smad2 is necessary for mesendoderm specification by Nodal signaling

A) Wild-type embryo at 36hpf B) MZoep embryo: maternal-zygotic mutant for one-eyed pinhead (*oep*) (Gritsman et al., 1999). C) MZ*smad2* embryo
D) MZ*smad2* embryo rescued with the injection of 20pg of *smad2* mRNA.

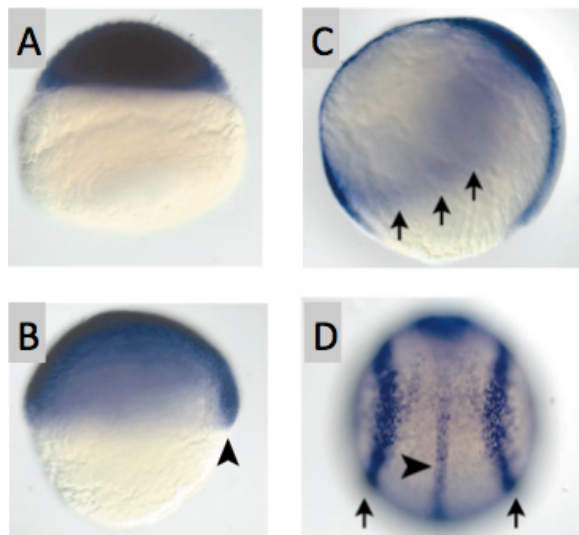
All R-Smads, as well as Smad4, can directly bind to DNA, Smad2 being the only exception (Massague et al., 2005). Smad3 and Smad4 have relatively weak DNA-binding affinity and their preferred binding site has the short motif GTCT (Zawel et al., 1998), which is commonly known as a Smad binding element (SBE). Many Nodal-regulated genes have one or more SBEs in the sequence of their enhancers (Ross and Hill, 2008). Since Smad2 doesn't have any DNA-binding ability, it requires additional sequence-specific transcription factors to recruit it to the regulatory loci (Massague et al., 2005). A number of such transcription factors have been identified, such as FoxH1, members of the homeodomain Mix/Bix transcription factor family (Mixer, Mezzo), eomesA, HEB, E2A, Oct1, and others (Kikuchi et al., 2000; Kunwar et al., 2003; Liu et al., 2011; Nelson et al., 2014; Poulain and Lepage, 2002; Yoon et al., 2011). Many of such transcription factors possess a conserved Smad-interacting domain (SID), mediating direct binding between these proteins and Smad2/3.

In order to find bona fide targets of Nodal signaling, it is crucial to identify Smad2 binding sites under different exposure conditions. In Chapter III I will describe the identification of Smad2 binding sites upon exposure to high levels of Nodal signaling, and testing these sites for their functionality.

Role of FoxH1 transcription factor in embryonic development

FoxH1 factor is one of the most important Smad2 co-factors, regulating a large number of Nodal-responsive genes in metazoans (Attisano et al., 2001). In the mouse, *FoxH1* is uniformly expressed in the embryo before and during gastrulation, then in the heart up to embryonic day E8.5, and subsequently disappears (Weisberg et al., 1998). Mouse *Foxh1* mutants display defects in the anterior-

posterior axis formation, primitive streak patterning, and anterior heart-field development (Hoodless et al., 2001; Yamamoto et al., 2001). In zebrafish embryos, *Foxh1* mRNA is expressed both maternally and zygotically (Sirotkin et al., 2000). Early on *FoxH1* transcripts are evenly distributed, until the beginning of gastrulation, when *FoxH1* mRNA is expressed in a ventral to dorsal gradient, with high levels on the ventral side and in the shield [Fig.1.5]. The expression of *FoxH1* becomes progressively restricted to midline and ventral cells, later on being detected exclusively in notochord, lateral plate mesoderm and in a stripe of anterior dorsal neuroectoderm [Fig.1.5] (Sirotkin et al., 2000). FoxH1 protein is not detected after the first day of development. Similar to *FoxH1*-deficient mouse embryos, zebrafish *FoxH1* mutants and *FoxH1*-depleted frog embryos display a loss of organizer and axial mesendoderm structures, indicating an evolutionarily conserved role for FoxH1 (Sirotkin et al., 2000; Slagle et al., 2011).



from (Sirotkin et al., 2000)

Figure 1.5

FoxH1 expression in early zebrafish embryo

A) At sphere stage *FoxH1* expression is ubiquitous.

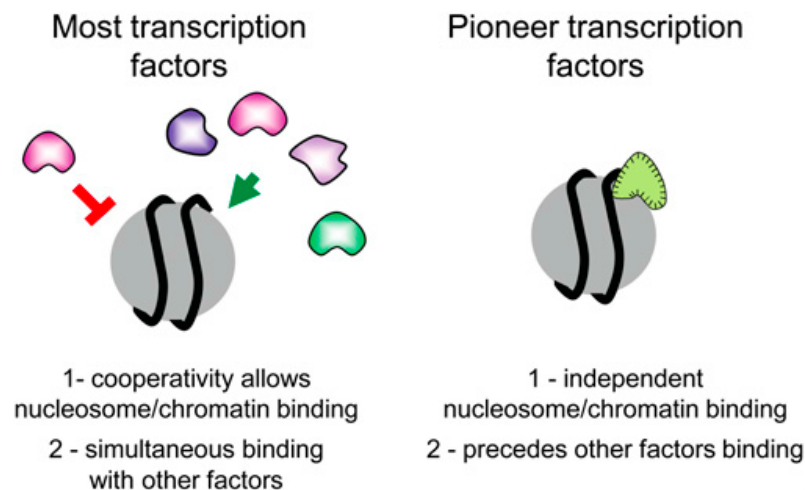
B) At 60% epiboly the gene is expressed in the shield (arrowhead) and in a ventral–dorsal gradient.

C) At 90% epiboly the gene is expressed in axial and lateral mesoderm (arrows).

D) At the one-somite stage the gene is expressed in notochord (arrowhead), lateral mesoderm (arrows) and the brain.

Pioneer factors in development and in signaling responses

Most of the nuclear DNA in eukaryotic organisms is inaccessible to transcription factors. Their genome is condensed into tightly packed chromatin, first by the wrapping of DNA around an octamer of histones to create the nucleosome core particle (Kornberg, 1974), then by the binding of linker histone near the nucleosome dyad axis (Zhou et al., 1998) and the folding of nucleosomes into higher order structures (Schwarz and Hansen, 1994). Indeed, genome-wide analysis of the binding of eukaryotic transcription factors always shows that only a small fraction of potential binding sites are occupied. One of the ways in which DNA stays or becomes accessible to transcription factors is the binding of ‘pioneer factors’ (Zaret and Carroll, 2011). This is a special class of proteins that let genes stay competent for expression, by accessing their DNA target sites even in compacted chromatin and keeping the site accessible for transcriptional activation. Unlike most transcription factors, which require cooperativity for binding to their target locus, pioneer factors can bind to DNA independent of other factors, and in some cases can recruit histone modifying enzymes to make the locus available for other proteins [Fig.1.6]. The action of pioneer factors is especially crucial for the regulation of processes that require a very rapid response, such as embryonic development or the activation of hormonal target genes. In the presence of a pioneer factor the enhancer is already marked, giving it a head start towards assembling the transcription preinitiation complex (Smale, 2010).



from (Zaret and Carroll, 2011)

Figure 1.6

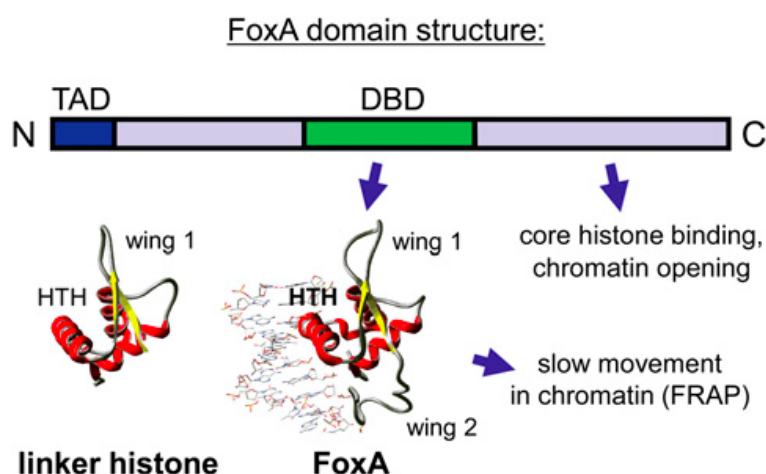
Characteristics of a pioneer factor

While most transcription factors require cooperativity for successful binding, pioneer transcription factors can precede the binding of other factors, and can bind chromatin independently.

The first protein that was called a pioneer factor was FoxA, which is involved in the development of endodermal tissues, enabling liver gene induction under the appropriate developmental signals (Cirillo et al., 2002). Some other known pioneer factors in mice and other vertebrates are PBX1, TLE, AP2 γ , GATA factors 2/3/4, and PU.1. Misregulation of pioneer factor expression often plays a role in the development of various types of cancer, making the identification and study of these factors all the more important (Jozwik and Carroll, 2012).

A number of other fork head proteins in addition to FoxA have been recognized as pioneer factors, including FoxD3, FoxE1, FoxI1, FoxO3, and FoxM (Cuesta et al., 2007; Xu et al., 2007; Yan et al., 2006; Yang and Hung, 2009). A common trait of fork head factors is that they contain a winged helix DNA-binding domain that mimics the DNA-binding domain of the linker H1 histone, allowing it to bind chromatin by interacting with the major groove of only the one available side of DNA wrapped around a nucleosome (Clark et al., 1993; Zaret et al., 2010). However,

unlike linker histone proteins, fork head domains also possess a helix in their structure, allowing for sequence specificity and higher mobility around the nucleosome (Sekiya et al., 2009). The active re-arrangement of the nucleosomes allows for other transcription factors to bind the available DNA [Fig.1.7].



from (Zaret and Carroll, 2011)

Figure 1.7

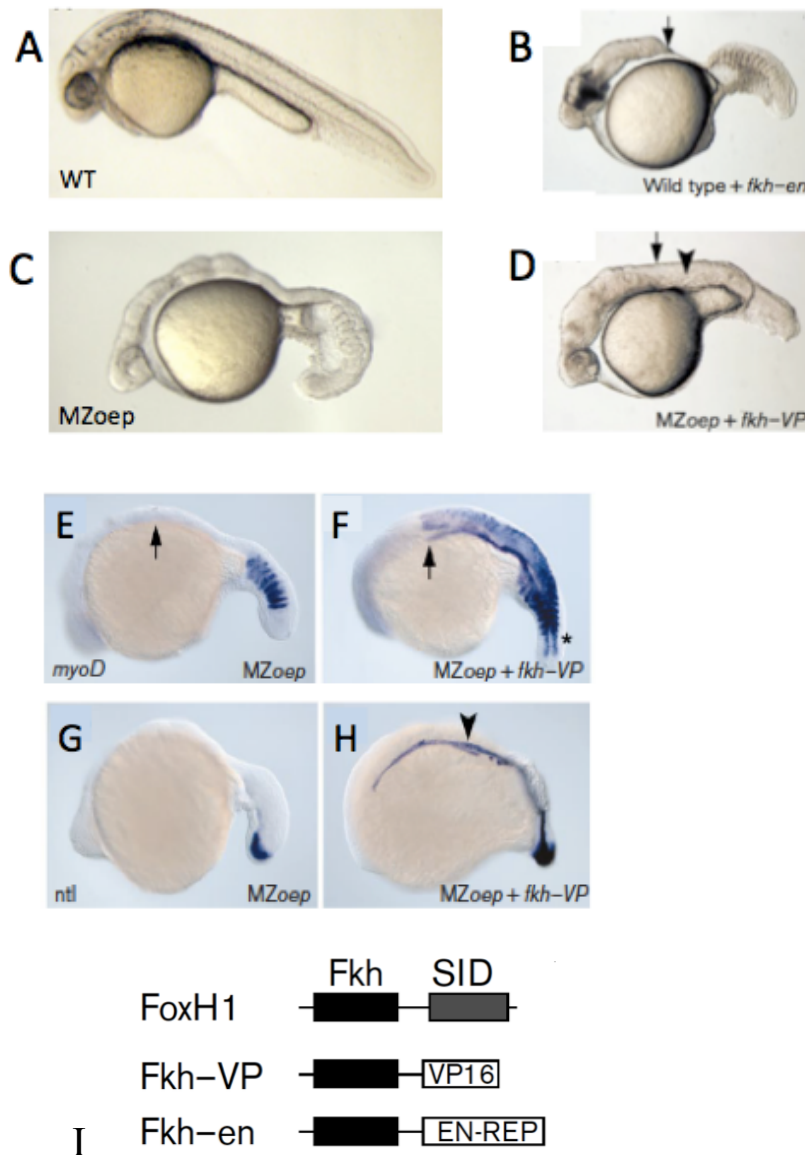
Forkhead factors possess structural features important for their pioneer factor role

The crystal structures of the DNA-binding domains of linker histone (Ramakrishnan et al., 1993) and FoxA3 (Clark et al., 1993) are shown side by side. The winged helix motif makes minor groove contacts along the long axis of the DNA.

Unlike linker histone, FoxA proteins make specific base contacts that direct target site binding and possess an N-terminal trans-activation domain (TAD) (Pani et al., 1992). FoxA proteins also possess a C-terminal domain that binds directly to core histone proteins and is necessary for the factor to open chromatin (Cirillo et al., 2002).

FoxH1 is an attractive candidate for the role of a pioneer factor in the Nodal signaling pathway. It is a fork head family protein, it is provided maternally, and there is a very considerable difference between the zygotic and maternal-zygotic mutant phenotype, showing that early presence of FoxH1 is crucial for the development. Previous experiments also show that the DNA-binding domain of FoxH1 is capable of binding to DNA independently (Pogoda et al., 2000). A series of experiments combining the DNA-binding FoxH1 domain with an activating (VP16)

or a repressing (Engrailed) domain showed that these constructs could activate or repress Nodal target genes in the absence of Nodal signals [Fig.1.8]. These observations raised a number of questions about the role of FoxH1 in zebrafish development. Can the full FoxH1 protein bind to nuclear DNA in the absence of Nodal? If so, is Smad2 recruitment its only function, or does it have Nodal-independent targets? I will address these questions in Chapter II, by analyzing gene expression in FoxH1 mutants and identifying FoxH1 binding sites in the presence and absence of Nodal signaling.



from (Pogoda et al., 2000)

Figure 1. 8

DNA-binding domain of FoxH1 is able to bind to DNA independent of Nodal signaling

A) Wild-type embryo B) Wild-type embryo injected with 10pg of *fkh-en* RNA developed a phenotype similar to *MZoep* mutants, including lack of trunk mesoderm and cyclopia (arrows mark the head-trunk boundary).

C) *MZoep* embryo D) Injection of 5pg of *fkh-VP* RNA partially rescued the formation of somitic mesoderm (arrow) and notochord (arrowhead) in *MZoep* mutants.

E-F) Rescued muscle-specific expression of *myoD* in trunk mesoderm in *fkh-VP* injected *MZoep* mutants,

G-H) Rescued expression of *ntla* in *fkh-VP* injected *MZoep* mutants.

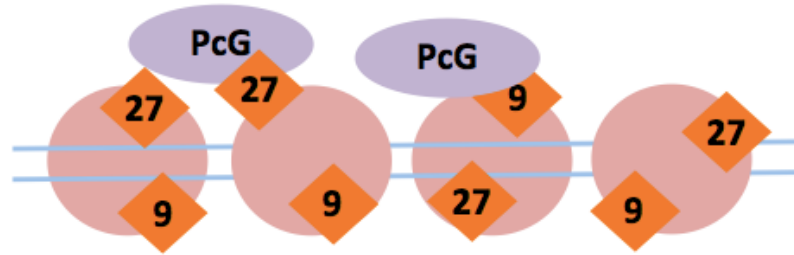
I) Constructs used for the experiment: wild-type FoxH1 protein; FoxH1 protein where Smad-interacting domain is replaced with VP16 domain; FoxH1 protein where Smad-interacting domain is replaced with the Engrailed domain.

The role of epigenetic marks in development and in signaling responses

While transcription factors can directly affect gene expression by interacting with the general transcription apparatus, they can also recruit various enzymes that modify the surrounding chromatin. These enzymes can rearrange the nucleosome positioning or modify the histones and DNA itself, leading to activation or repression of gene expression (Li et al., 2007). Smad2 and Smad3, for example, have been shown to recruit co-activator p300, which is a histone acetyltransferase, to gene regulatory elements (Feng et al., 1998; Janknecht et al., 1998). Smad2 interacts with a number of other histone-modifying enzymes to modulate the expression of Nodal target genes (Ross and Hill, 2008).

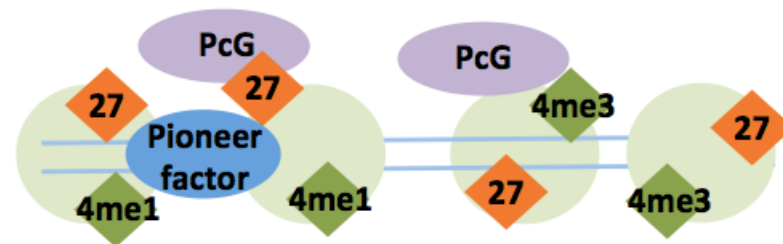
Numerous mapping studies have detected correlations between the chromatin state of a gene and its levels of expression (Fisher and Fisher, 2011; Siggens and Ekwall, 2014), and knockout experiments show that loss of histone-modifying proteins leads to developmental abnormalities. For example, the Polycomb group proteins in the PRC2 complex di- and tri-methylate H3K27 and are required for embryo viability and normal gene expression during embryonic stem (ES) cell differentiation (Erhardt et al., 2003; Faust et al., 1998, 1995).

Enhancer elements experience a change in chromatin structure upon being primed for and the actual activation of the gene it regulates. H3K9me3 mark is associated with silent enhancers, H3K27me3 – with silent or poised ones, and H3K4me1 with poised enhancers only (Bonn et al., 2012; Rada-Iglesias et al., 2011; Zhu et al., 2013, 2012). On the other hand, acetylation of H3K27 residue (H3K27ac) is a mark of active enhancers (Creyghton et al., 2010) [Fig.1.9].



INACTIVE STATE

Repressed chromatin is associated with the presence of H3K27me3 and H3K9me3 marks, binding of Polycomb group (PcG) proteins, DNA methylation and compact nucleosome positioning.

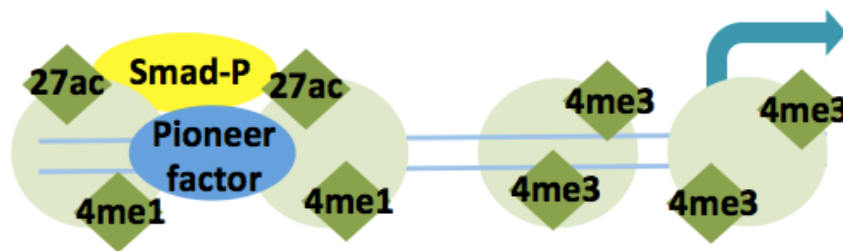


POISED STATE

Enhancer

Promoter

Poised enhancers and promoters still have H3K27me3 histone marks and PcG proteins bound on them, but there is an overall less compact chromatin structure and lack of DNA methylation. Poised promoters are associated with H3K4me3, and poised enhancer with H3K4me1 histone marks. Pioneer factors may be bound to poised enhancers.



ACTIVE STATE

Enhancer

Promoter

Active promoters lose H3K27me3 mark. Active enhancers lose H3K27me3 modification and are marked by H3K4me1 and H3K27ac.

Figure 1.9

Chromatin structure and its modifications reflect and govern the transcriptional state of the associated genes

A number of marks exhibit very strong association with either silenced (e.g. H3K27me3) or active (e.g. H3K4me3) gene promoters (Rando, 2007). However, some genes contain both activating and repressing marks at their promoters (combination of H3K27me and H3K4me3), and it has been proposed that this combination of marks keeps genes turned off but poised for activation (Bernstein et al., 2006; Vastenhouw and Schier, 2012). Important developmental genes, such as genes activated by patterning signals, often possess such bivalent marks (Bernstein et al., 2006; Vastenhouw and Schier, 2012). Studies in cell culture show that many Nodal-regulated genes are bivalent in embryonic stem cells, and the bivalent mark resolves upon activation of the genes. The levels of the silencing H3K27me3 mark are reduced and levels of H3K4me3 mark go up in response to Nodal exposure (Kim et al., 2011) [Fig.1.9].

How Nodal signaling affects chromatin affects chromatin landscape of its target genes has not yet been studied in a developing embryo, and it is unclear whether the process is identical to that in cell culture. It is also unknown whether simply getting rid of a silencing mark is sufficient to change the expression of a Nodal target gene. I will address these questions in Chapter IV, as I study the distribution of H3K4me3 and H3K27me3 marks in embryos with and without Nodal signal exposure, and test the effect of targeted H3K27me3 mark reduction on the expression of a Nodal target gene.

Overview

The regulation of transcriptional response to signaling is important for our understanding of all areas of biology, from basic embryonic development to biomedical applications. A great deal of research has been done in the area in the past several decades, but there are still many questions to be answered. In particular, as many studies are done in cell culture, it is often unclear whether these findings hold true in an actual developing organism. Experiments done in zebrafish, a vertebrate fast-developing organism, thus become especially important. As misregulation of Nodal signaling is implicated in tumorigenesis and metastasis in a number of cancer types, studying this signal has biomedical significance.

My research described in the following chapters concerns the transcriptional regulation of Nodal target genes in an early zebrafish embryo. I studied the process of transcription factor assembly on Nodal regulated enhancers, focusing on FoxH1 and Smad2 factors, and the chromatin modification changes at the promoters of Nodal target genes. In Chapter II I study the role that FoxH1 plays in the regulation of Nodal target genes, as well as its potential Nodal-independent target genes. Chapter III describes the identification of Smad2 binding sites upon oversaturation of the system with Nodal signaling and discovery of novel Nodal target genes. In Chapter IV I describe the chromatin state of Nodal target genes in the presence and absence of Nodal signaling, and the use of Cas9-based tools for targeted reduction of H3K27me3 mark.

REFERENCES

- Attisano, L., Silvestri, C., Izzi, L., Labbé, E., 2001. The transcriptional role of Smads and FAST (FoxH1) in TGF β and activin signalling, in: *Molecular and Cellular Endocrinology*. pp. 3–11. doi:10.1016/S0303-7207(01)00524-X
- Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S.L., Lander, E.S., 2006. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* 125, 315–326. doi:10.1016/j.cell.2006.02.041
- Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczyński, B., Riddell, A., Furlong, E.E.M., 2012. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat. Genet.* 44, 148–156. doi:10.1038/ng.1064
- Chu, J., Ding, J., Jeays-Ward, K., Price, S.M., Placzek, M., Shen, M.M., 2005. Non-cell-autonomous role for Cripto in axial midline formation during vertebrate embryogenesis. *Development* 132, 5539–51. doi:10.1242/dev.02157
- Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., Zaret, K.S., 2002. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* 9, 279–289. doi:10.1016/S1097-2765(02)00459-8
- Clark, K.L., Halay, E.D., Lai, E., Burley, S.K., 1993. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* 364, 412–420. doi:10.1038/364412a0
- Conlon, F.L., Lyons, K.M., Takaesu, N., Barth, K.S., Kispert, a, Herrmann, B., Robertson, E.J., 1994. A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. *Development* 120, 1919–1928. doi:10.1016/0168-9525(94)90026-4
- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., Boyer, L.A., Young, R.A., Jaenisch, R., 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* 107, 21931–21936. doi:10.1073/pnas.1016071107
- Cuesta, I., Zaret, K.S., Santisteban, P., 2007. The forkhead factor FoxE1 binds to the thyroperoxidase promoter during thyroid cell differentiation and modifies compacted chromatin structure. *Mol. Cell. Biol.* 27, 7302–7314.

doi:10.1128/MCB.00758-07

- Dawson, M.A., Kouzarides, T., 2012. Cancer epigenetics: From mechanism to therapy. *Cell*. doi:10.1016/j.cell.2012.06.013
- Driever, W., Nusslein-Volhard, C., 1988. The bicoid protein determines position in the *Drosophila* embryo in a concentration-dependent manner. *Cell* 54, 95–104. doi:10.1016/0092-8674(88)90183-3
- Dubrulle, J., Jordan, B.M., Akhmetova, L., Farrell, J.A., Kim, S.-H., Solnica-Krezel, L., Schier, A.F., 2015. Response to Nodal morphogen gradient is determined by the kinetics of target gene induction. *Elife* 4. doi:10.7554/eLife.05042
- Erhardt, S., Su, I.-H., Schneider, R., Barton, S., Bannister, A.J., Perez-Burgos, L., Jenuwein, T., Kouzarides, T., Tarakhovsky, A., Surani, M.A., 2003. Consequences of the depletion of zygotic and embryonic enhancer of zeste 2 during preimplantation mouse development. *Development* 130, 4235–4248. doi:10.1242/dev.00625
- Erter, C.E., Solnica-Krezel, L., Wright, C. V, 1998. Zebrafish nodal-related 2 encodes an early mesendodermal inducer signaling from the extraembryonic yolk syncytial layer. *Dev. Biol.* 204, 361–372. doi:10.1006/dbio.1998.9097
- Faust, C., Lawson, K.A., Schork, N.J., Thiel, B., Magnuson, T., 1998. The Polycomb-group gene *eed* is required for normal morphogenetic movements during gastrulation in the mouse embryo. *Development* 125, 4495–4506.
- Faust, C., Schumacher, A., Holdener, B., Magnuson, T., 1995. The *eed* mutation disrupts anterior mesoderm production in mice. *Development* 121, 273–285.
- Feldman, B., Gates, M.A., Egan, E.S., Dougan, S.T., Rennebeck, G., Sirotkin, H.I., Schier, A.F., Talbot, W.S., 1998a. Zebrafish organizer development and germ-layer formation require nodal-related signals. *Nature* 395, 181–185. doi:10.1038/26013
- Feldman, B., Gates, M.A., Egan, E.S., Dougan, S.T., Rennebeck, G., Sirotkin, H.I., Schier, A.F., Talbot, W.S., 1998b. Zebrafish organizer development and germ-layer formation require nodal-related signals. *Nature* 395, 181–5. doi:10.1038/26013
- Feng, X.H., Zhang, Y., Wu, R.Y., Derynck, R., 1998. The tumor suppressor Smad4/DPC4 and transcriptional adaptor CBP/p300 are coactivators for Smad3 in TGF- β -induced transcriptional activation. *Genes Dev.* 12, 2153–2163. doi:10.1101/gad.12.14.2153
- Fisher, C.L., Fisher, A.G., 2011. Chromatin states in pluripotent, differentiated, and reprogrammed cells. *Curr. Opin. Genet. Dev.*
- Gierer, A., Meinhardt, H., 1972. A Theory of Biological Pattern Formation.

- Kybernetik 12, 30–39. doi:10.1007/BF00289234
- Green, J.B.A., Sharpe, J., 2015. Positional information and reaction-diffusion: two big ideas in developmental biology combine. *Development* 142, 1203–1211. doi:10.1242/dev.114991
- Gritsman, K., Zhang, J., Cheng, S., Heckscher, E., Talbot, W.S., Schier, a F., 1999. The EGF-CFC protein one-eyed pinhead is essential for nodal signaling. *Cell* 97, 121–132. doi:10.1016/S0092-8674(00)80720-5
- Hoodless, P.A., Pye, M., Chazaud, C., Labbé, E., Attisano, L., Rossant, J., Wrana, J.L., 2001. FoxH1 (Fast) functions to specify the anterior primitive streak in the mouse. *Genes Dev.* 15, 1257–1271. doi:10.1101/gad.881501
- Janknecht, R., Wells, N.J., Hunter, T., 1998. TGF- β -stimulated cooperation of Smad proteins with the coactivators CBP/p300. *Genes Dev.* 12, 2114–2119. doi:10.1101/gad.12.14.2114
- Jozwik, K.M., Carroll, J.S., 2012. Pioneer factors in hormone-dependent cancers. *Nat. Rev. Cancer* 12, 381–385. doi:10.1038/nrc3263
- Kikuchi, Y., Trinh, L.A., Reiter, J.F., Alexander, J., Yelon, D., Stainier, D.Y.R., 2000. The zebrafish bonnie and clyde gene encodes a Mix family homeodomain protein that regulates the generation of endodermal precursors. *Genes Dev.* 14, 1279–1289. doi:10.1101/gad.14.10.1279
- Kim, S.W., Yoon, S.J., Chuong, E., Oyolu, C., Wills, A.E., Gupta, R., Baker, J., 2011. Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. *Dev. Biol.* 357, 492–504.
- Kornberg, R.D., 1974. Chromatin structure: a repeating unit of histones and DNA. *Science* 184, 868–871. doi:10.1126/science.184.4139.868
- Kunwar, P.S., Zimmerman, S., Bennett, J.T., Chen, Y., Whitman, M., Schier, A.F., 2003. Mixer/Bon and FoxH1/Sur have overlapping and divergent roles in Nodal signaling and mesendoderm induction. *Development* 130, 5589–5599. doi:10.1242/dev.00803
- Le Good, J.A., Joubin, K., Giraldez, A.J., Ben-Haim, N., Beck, S., Chen, Y., Schier, A.F., Constam, D.B., 2005. Nodal stability determines signaling range. *Curr. Biol.* 15, 31–36. doi:10.1016/j.cub.2004.12.062
- Li, B., Carey, M., Workman, J.L., 2007. The Role of Chromatin during Transcription. *Cell*. doi:10.1016/j.cell.2007.01.015
- Lindroth, A.M., Park, Y.J., Plass, C., 2015. Epigenetic reprogramming in cancer. *Epigenetics Hum. Heal.* 4, 193–223. doi:10.1007/978-3-642-31974-7_9
- Liu, Z., Lin, X., Cai, Z., Zhang, Z., Han, C., Jia, S., Meng, A., Wang, Q., 2011.

- Global identification of SMAD2 target genes reveals a role for multiple co-regulatory factors in zebrafish early gastrulas. *J. Biol. Chem.* 286, 28520–28532. doi:10.1074/jbc.M111.236307
- Massague, J., 2012. TGF beta signalling in context. *Nat. Rev. Mol. CELL Biol.* 13, 616–630. doi:10.1038/nrm3434
- Massague, J., Seoane, J., Wotton, D., 2005. Smad transcription factors. *Genes Dev.* doi:10.1101/gad.1350705
- Muller, P., Rogers, K.W., Jordan, B.M., Lee, J.S., Robson, D., Ramanathan, S., Schier, a. F., 2012. Differential Diffusivity of Nodal and Lefty Underlies a Reaction-Diffusion Patterning System. *Science* (80-.). 336, 721–724. doi:10.1126/science.1221920
- Nelson, A.C., Cutty, S.J., Niini, M., Stemple, D.L., Flicek, P., Houart, C., Bruce, A.E.E., Wardle, F.C., 2014. Global identification of Smad2 and Eomesodermin targets in zebrafish identifies a conserved transcriptional network in mesendoderm and a novel role for Eomesodermin in repression of ectodermal gene expression. *BMC Biol.* 12, 81. doi:10.1186/s12915-014-0081-5
- Oshimori, N., Fuchs, E., 2012. The harmonies played by TGF- β in stem cell biology. *Cell Stem Cell.* doi:10.1016/j.stem.2012.11.001
- Pang, K., Ryan, J.F., Baxeavanis, A.D., Martindale, M.Q., 2011. Evolution of the TGF-?? signaling pathway and its potential role in the ctenophore, mnemiopsis leidyi. *PLoS One* 6. doi:10.1371/journal.pone.0024152
- Pani, L., Overdier, D.G., Porcella, a, Qian, X., Lai, E., Costa, R.H., 1992. Hepatocyte nuclear factor 3 beta contains two transcriptional activation domains, one of which is novel and conserved with the Drosophila fork head protein. *Mol. Cell. Biol.* 12, 3723–32. doi:10.1128/MCB.12.9.3723.Updated
- Pogoda, H.M., Solnica-Krezel, L., Driever, W., Meyer, D., 2000. The zebrafish forkhead transcription factor FoxH1/Fast1 is a modulator of Nodal signaling required for organizer formation. *Curr. Biol.* 10, 1041–1049. doi:10.1016/S0960-9822(00)00669-2
- Poulain, M., Lepage, T., 2002. Mezzo, a paired-like homeobox protein is an immediate target of Nodal signalling and regulates endoderm specification in zebrafish. *Development* 129, 4901–4914.
- Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. a, Flynn, R. a, Wysocka, J., 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470, 279–83. doi:10.1038/nature09692
- Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L., Sweet, R.M., 1993. Crystal structure of globular domain of histone H5 and its implications for nucleosome

- binding. *Nature* 362, 219–223. doi:10.1038/362219a0
- Rando, O.J., 2007. Global patterns of histone modifications. *Curr. Opin. Genet. Dev.* 17, 94–99. doi:10.1016/j.gde.2007.02.006
- Rebagliati, M.R., Toyama, R., Fricke, C., Haffter, P., Dawid, I.B., 1998. Zebrafish nodal-related genes are implicated in axial patterning and establishing left-right asymmetry. *Dev. Biol.* 199, 261–272. doi:10.1006/dbio.1998.8935
- Ross, S., Hill, C.S., 2008. How the Smads regulate transcription. *Int. J. Biochem. Cell Biol.* doi:10.1016/j.biocel.2007.09.006
- Sampath, K., Rubinstein, a L., Cheng, a M., Liang, J.O., Fekany, K., Solnica-Krezel, L., Korzh, V., Halpern, M.E., Wright, C. V, 1998. Induction of the zebrafish ventral brain and floorplate requires cyclops/nodal signalling. *Nature* 395, 185–189. doi:10.1038/26020
- Schier, A.F., 2009. Nodal morphogens. *Cold Spring Harb. Perspect. Biol.* doi:10.1101/cshperspect.a003459
- Schier, A.F., Shen, M.M., 2000. Nodal signalling in vertebrate development. *Nature* 403, 385–389. doi:10.1038/35000126
- Schwarz, P.M., Hansen, J.C., 1994. Formation and stability of higher order chromatin structures Contributions of the histone octamer. *J. Biol. Chem.* 269, 16284–16289.
- Sekiya, T., Muthurajan, U.M., Luger, K., Tulin, A. V., Zaret, K.S., 2009. Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. *Genes Dev.* 23, 804–809. doi:10.1101/gad.1775509
- Shen, M.M., Schier, A.F., 2000. The EGF-CFC gene family in vertebrate development. *Trends Genet.* doi:10.1016/S0168-9525(00)02006-0
- Siggens, L., Ekwall, K., 2014. Epigenetics, chromatin and genome organization: Recent advances from the ENCODE project. *J. Intern. Med.* doi:10.1111/joim.12231
- Sirotkin, H.I., Gates, M.A., Kelly, P.D., Schier, A.F., Talbot, W.S., 2000. fast1 is required for the development of dorsal axial structures in zebrafish. *Curr. Biol.* 10, 1051–1054. doi:10.1016/S0960-9822(00)00679-5
- Slagle, C.E., Aoki, T., Burdine, R.D., 2011. Nodal-dependent mesendoderm specification requires the combinatorial activities of FoxH1 and eomesodermin. *PLoS Genet.* 7. doi:10.1371/journal.pgen.1002072
- Smale, S.T., 2010. Pioneer factors in embryonic stem cells and differentiation. *Curr. Opin. Genet. Dev.* 20, 519–26. doi:10.1016/j.gde.2010.06.010

- Strizzi, L., Hardy, K.M., Kirschmann, D.A., Ahrlund-Richter, L., Hendrix, M.J.C., 2012. Nodal expression and detection in cancer: Experience and challenges. *Cancer Res.* doi:10.1158/0008-5472.CAN-11-3419
- Tsuchida, K., Nakatani, M., Yamakawa, N., Hashimoto, O., Hasegawa, Y., Sugino, H., 2004. Activin isoforms signal through type I receptor serine/threonine kinase ALK7. *Mol. Cell. Endocrinol.* 220, 59–65. doi:10.1016/j.mce.2004.03.009
- Turing, a M., 1952. The Chemical Basis of Morphogenesis THE CHEMICAL BASIS OF MOKPHOGENESIS. *Society* 237, 37–72. doi:10.1098/rstb.1952.0012
- van Boxtel, A.L., Chesebro, J.E., Heliot, C., Ramel, M.C., Stone, R.K., Hill, C.S., 2015. A Temporal Window for Signal Activation Dictates the Dimensions of a Nodal Signaling Domain. *Dev. Cell* 35, 175–185. doi:10.1016/j.devcel.2015.09.014
- Vastenhouw, N.L., Schier, A.F., 2012. Bivalent histone modifications in early embryogenesis. *Curr. Opin. Cell Biol.* 24, 374–86. doi:10.1016/j.ceb.2012.03.009
- Weisberg, E., Winnier, G.E., Chen, X., Farnsworth, C.L., Hogan, B.L.H., Whitman, M., 1998. A mouse homologue of FAST-1 transduces TGF?? superfamily signals and is expressed during early embryogenesis. *Mech. Dev.* 79, 17–27. doi:10.1016/S0925-4773(98)00160-9
- Wolpert, L., 1969. Positional information and the spatial pattern of cellular differentiation. *J. Theor. Biol.* 25, 1–47. doi:10.1016/S0022-5193(69)80016-0
- Wrana, J.L., Attisano, L., Wieser, R., Ventura, F., Massague, J., 1994. Mechanism of activation of the TGF- β receptor. *Nature* 370, 341–347. doi:10.1038/370341a0
- Xu, J., Pope, S.D., Jazirehi, A.R., Attema, J.L., Papathanasiou, P., Watts, J. a, Zaret, K.S., Weissman, I.L., Smale, S.T., 2007. Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 104, 12377–12382. doi:10.1073/pnas.0704579104
- Yamamoto, M., Meno, C., Sakai, Y., Shiratori, H., Mochida, K., Ikawa, Y., Saijoh, Y., Hamada, H., 2001. The transcription factor FoxH1 (FAST) mediates Nodal signaling during anterior-posterior patterning and node formation in the mouse. *Genes Dev.* 15, 1242–1256. doi:10.1101/gad.883901
- Yan, J., Xu, L., Crawford, G., Wang, Z., Burgess, S.M., 2006. The forkhead transcription factor FoxI1 remains bound to condensed mitotic chromosomes and stably remodels chromatin structure. *Mol. Cell. Biol.* 26, 155–68. doi:10.1128/MCB.26.1.155-168.2006

- Yang, J.Y., Hung, M.C., 2009. A new fork for clinical application: Targeting forkhead transcription factors in cancer. *Clin. Cancer Res.* doi:10.1158/1078-0432.CCR-08-0124
- Yoon, S.J., Wills, A.E., Chuong, E., Gupta, R., Baker, J.C., 2011. HEB and E2A function as SMAD/FOXH1 cofactors. *Genes Dev.* 25, 1654–1661. doi:10.1101/gad.16800511
- Zaret, K.S., Caravaca, J.M., Tulin, A., Sekiya, T., 2010. Nuclear mobility and mitotic chromosome binding similarities between pioneer transcription factor FoxA and linker histone H1. *Cold Spring Harb. Symp. Quant. Biol.* 75, 219–226. doi:10.1101/sqb.2010.75.061
- Zaret, K.S., Carroll, J.S., 2011. Pioneer transcription factors: Establishing competence for gene expression. *Genes Dev.* doi:10.1101/gad.176826.111
- Zawel, L., Dai, J.L., Buckhaults, P., Zhou, S., Kinzler, K.W., Vogelstein, B., Kern, S.E., 1998. Human Smad3 and Smad4 are sequence-specific transcription activators. *Mol. Cell* 1, 611–7. doi:10.1016/S1097-2765(00)80061-1
- Zhou, X., Sasaki, H., Lowe, L., Hogan, B.L., Kuehn, M.R., 1993. Nodal is a novel TGF-beta-like gene expressed in the mouse node during gastrulation. *Nature* 361, 543–547. doi:10.1038/361543a0
- Zhou, Y.B., Gerchman, S.E., Ramakrishnan, V., Travers, a, Muylldermans, S., 1998. Position and orientation of the globular domain of linker histone H5 on the nucleosome. *Nature* 395, 402–405. doi:10.1038/26521
- Zhu, Y., Sun, L., Chen, Z., Whitaker, J.W., Wang, T., Wang, W., 2013. Predicting enhancer transcription and activity from chromatin modifications. *Nucleic Acids Res.* 41, 10032–10043. doi:10.1093/nar/gkt826
- Zhu, Y., van Essen, D., Sacconi, S., 2012. Cell-Type-Specific Control of Enhancer Activity by H3K9 Trimethylation. *Mol. Cell* 46, 408–423. doi:10.1016/j.molcel.2012.05.011

CHAPTER II

Analysis of FoxH1 function and target genes

ABSTRACT

In this chapter I will talk about in depth analysis of FoxH1 function and its target genes. FoxH1 has been shown to play a range of roles, including Nodal-independent ones, in early vertebrate development in various model organisms (Chiu et al., 2014; Silvestri et al., 2008). I generated a complete mutant of this transcription factor and used deep sequencing to identify which genes FoxH1 regulates in zebrafish development. I did not detect a pre-gastrulation phenotype in these mutants, contrary to observations made in FoxH1 morphants (Pei et al., 2007). Using ChIP-Seq experiments, I also found the binding sites of FoxH1. I found that FoxH1 is capable of binding to genomic DNA in the absence of Nodal signaling, which suggests that it may act as a pioneer factor, preparing target genes for rapid activation when gastrulation starts. I also found 54 direct FoxH1 target genes that are not Nodal-dependent. Genes involved in nervous system development were overrepresented in this group.

PREFACE

I designed, performed, and interpreted all experiments. Summer Thyme generated the Cas9 protein and targeting RNA used to create the *FoxH1* mutants, and provided help during the mutant generation process. The Burdine lab generously provided the *midway* fish. Smad2 and FoxH1 ChIP-Seq data has been previously published in (Dubrulle et al., 2015). Alexander F. Schier helped conceive, design, and support the project.

INTRODUCTION

FoxH1 (Fast1, fork head activin signal transducer 1), is a winged helix transcription factors that was first discovered as a mediator of Activin-like signaling in conjunction with Smad2/3 (Chen et al., 1996). It has been well studied in cell culture and in *Xenopus* embryos, where it was shown to also function independently of Nodal signaling (Chiu et al., 2014; Kim et al., 2011; Silvestri et al., 2008). Two zebrafish mutants for the FoxH1 transcription factor have been created and analyzed so far. However, both of the published mutants, *schmalspur* (*sur*) (Kunwar et al., 2003) and *midway* (*mid*) (Slagle et al., 2011), do not get rid of the protein completely, but rather create a single amino-acid DNA-binding site mutation and a truncated version of the protein, respectively [Fig.2.1]. Comparison between *sur* and *mid* phenotypes suggested that *sur* is a hypomorphic allele. This was tested by injecting wild-type and mutant embryos with 50 pg of *squint* mRNA at one-cell stage, fixing them at 30–50% epiboly, and then visualizing the expression of *ntla* and *gsc* genes. While wild-type and MZ*sur* embryos upregulated both targets ubiquitously, MZ*mid* embryos exhibited a much weaker Nodal-dependent activation of both genes.

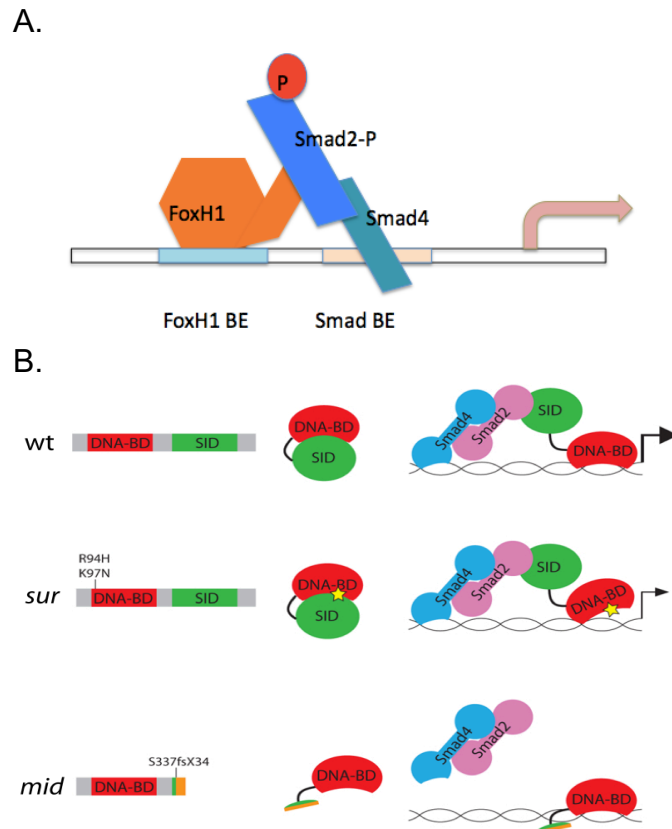


Figure 2.1

FoxH1 transcription factor in zebrafish and its published mutants

A) FoxH1 transcription factor binds to enhancer region of a regulated gene containing the consensus sequence {AATAMACA} and recruits activated phosphorylated Smad2 factors, which in their turn interact with Smad binding elements in a weak manner.

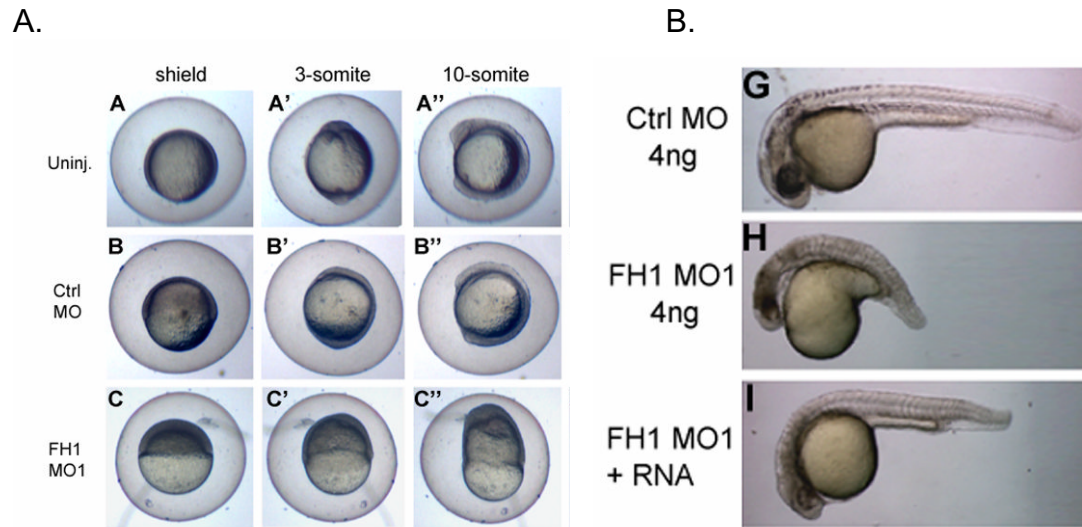
B) (from (Slagle et al., 2011)) Schematic representation of the wild type FoxH1 factor, with DNA-binding domain shown in red, and Smad-interacting domain (SID) shown in green.

sur mutant variant cannot independently bind to DNA because of a missense mutation in the DNA-binding domain, but could potentially be stabilized at a loci in the presence of activated Smad complexes.

mid mutant variant possesses a frame-shift mutation leading to a truncation right before SID, but the protein is not completely removed from the cell, leaving the possibility of residual activating or repressive ability.

As the DNA-binding domain of FoxH1 is unperturbed in *mid* mutants, there is a possibility that some function remains. Since in the *mid* fish Smad-binding domain of FoxH1 is removed, the remaining part of the protein might even act as a repressor of the genes that it normally activates. A study of zebrafish FoxH1 morphants (Pei et al., 2007), for example, suggests that FoxH1 has an early

requirement for maternal FoxH1 during gastrulation that is not observed in *mid* mutants [Fig.2.2]. We therefore created new mutant alleles of FoxH1, to completely get rid of this protein and all its domains. We then analyzed this mutant to fully understand the function of FoxH1 in early zebrafish development.



from (Pei et al., 2007)

Figure 2.2

FoxH1 morphant exhibits an early gastrulation phenotype

A) Injection of *FoxH1* translation-blocking morpholino, but not control morpholino, disrupts embryonic epiboly movements during gastrulation and causes death on the first day of development. Apart from disrupting epiboly, *FoxH1* MO treatment disrupts convergence and internalization movements.

Top to bottom: phenotype of an uninjected embryo, embryo injected with control MO, and FoxH1 translation-blocking MO.

Left to right: shield, 3-somite, and 10-somite developmental stages.

B) Morphant phenotype can be rescued by co-injection of *FoxH1* mRNA

Top to bottom: phenotype of embryo injected with control MO, injected with *FoxH1* MO and coinjected with *FoxH1* MO and *FoxH1* mRNA.

RESULTS

FoxH1 mutant generation and allele description

In order to generate a complete FoxH1 mutant, we used Cas9-mediated targeted mutagenesis. To make sure that all possible isoforms will be mutated, we targeted 3 different regions in the *foxH1* gene, one in each of the exons of the gene [Fig.2.3.A]. Multiple gRNAs were designed in order to target each of these regions. Summer Thyme performed the design of these gRNAs and the initial injection of those together with Cas9 protein at 1-cell stage of the embryos, as previously described (Gagnon et al., 2014). We obtained alleles containing frame-shift mutations at each of the targeted locations. Heterozygous embryos survived to adulthood and were crossed together to obtain homozygous fish.

A.



B.

foxh1e143	101	TCCACAGGGATGGCGCAGGACTCTTGTATAGAGCCAAAGCGACGAATCA	
I1_R_32			
	48	TCCACAGGGATGGCGCAGGACTCTTGTATAGAGCCAAAGCGACGAATCA	
foxh1e143	151	AGGCCCATGGGAAGTGCAGGATGGAACTCCAGTGGAGGAAAGAGAAAA	
I1_R_32			
	98	A-----GGGAAGTGCAGGATGGAACTCCAGTGGAGGAAAGAGAAAA	
foxh1e143	201	ACTATCAGCG	210
I1_R_32			
	141	ACTATCAGCG	150

C.

foxh1e316	51	CGTACATCTTTCAAGGTTACAGTCAGCCGAACAAAGTCCAAACCTCTTCCC
I0_F_333	51	CGTACATCTTTCAAGGTTACAGTCAGCCGAACAAAGTCC-----
foxh1e316	101	CCTGAATCATCTTTACCCCTGTTCACCCGTCACAGCCACCTCCATC
I0_F_333	89	-----AATCATCTTTACCCCTGTTCACCCGTCACAGCCACCTCCATC
foxh1e316	151	AGAGGACCCATATCGTCCCAATTGGACTCTACGTTTGCTATAGATTCTC
I0_F_333	135	AGAGGACCCATATCGT-----

Figure 2.3

FoxH1 mutant generation using Cas9-mediated mutagenesis

A) Targeted sites in *FoxH1* gene in the first, second and third exons.

B) Sequence alignment results showing a 7bp deletion in exon 1.

C) Sequence alignment results showing a 16bp deletion in exon 3.

For in depth analysis, we focused on two alleles, both with a frame-shift mutation: one a 7bp deletion in exon 1, another a 16bp deletion in exon 3 [Fig.2.3.B-C]. Zygotic homozygous fish with either allele exhibited defects in development, such as severe ventral body curvature, and did not survive past 4 days [Fig.2.4]. These homozygous fish could be rescued by injection of 10pg of *FoxH1* mRNA at 1-cell stage. The rescued embryos developed normally, and crossing them generated maternal-zygotic (MZ) mutant progeny. All analysis hereafter has been done using the allele with a frame-shift mutation in exon 1 of the gene (*E1* mutant).

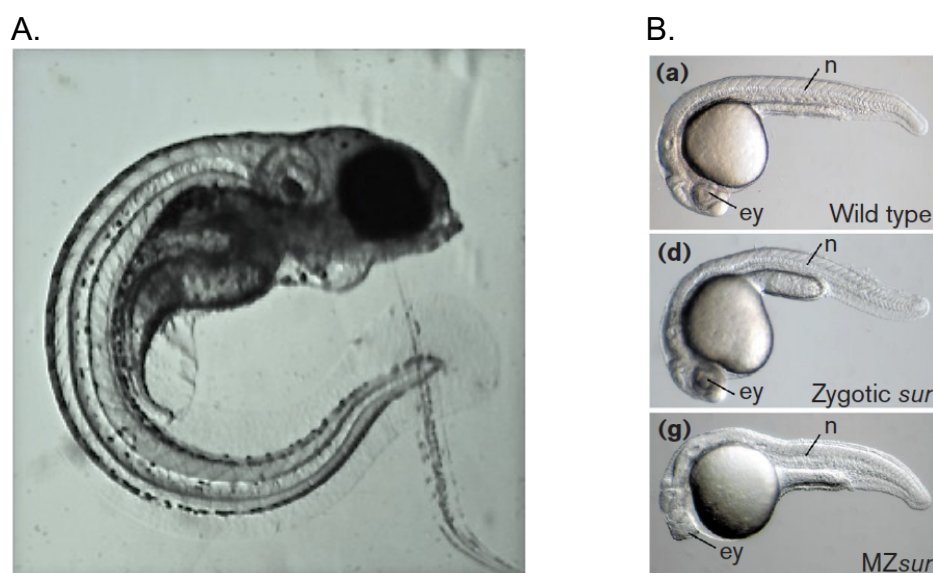


Figure 2.4:

Zygotic homozygous FoxH1 E1 mutant embryo

A) Zygotic homozygous *E1* mutant embryo at 48hpf. These embryos exhibit severe developmental abnormalities and do not survive past 4 days.

B) For comparison, wild type, zygotic *sur* and maternal-zygotic homozygous *sur* embryos are shown on the right (from (Kunwar et al., 2003)).

FoxH1 maternal-zygotic mutant phenotype

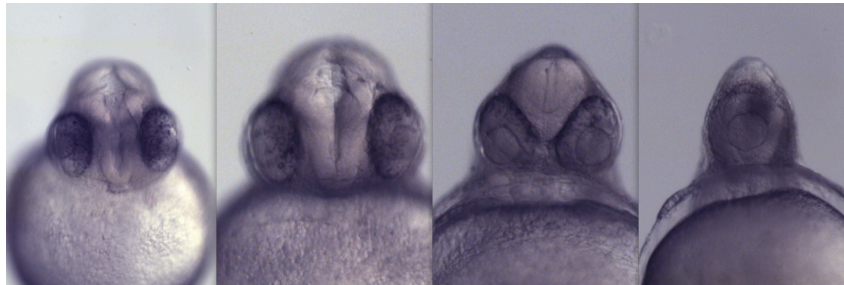
Contrary to what experiments with FoxH1 morpholino suggested, I observed no pre-gastrulation phenotype in MZ*E1* embryos (Pei et al., 2007). Embryos started exhibiting differences from wild type ones only at shield stage (6hpf). At

24hpf, these embryos consistently displayed an absence of notochord and very close-set eyes (often full cyclopia), hallmarks of Nodal signaling deficits [Fig.2.5].

A. wild-type || *E1* zygotic mutant || *E1* maternal-zygotic mutants



B. wild-type || *E1* zygotic mutant || *E1* maternal-zygotic mutants



C. wild-type || *E1* zygotic mutant || *E1* maternal-zygotic mutants

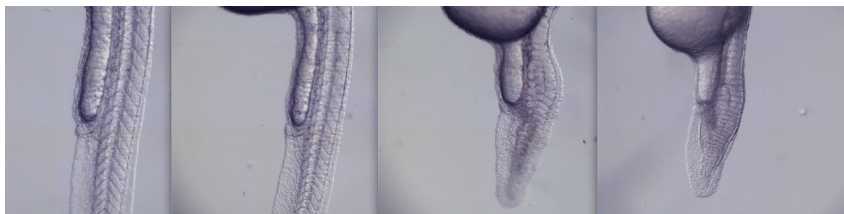


Figure 2.5

FoxH1 maternal-zygotic *E1* mutant phenotype

A) Whole embryo view: shortened trunk and underdeveloped head are clearly visible.

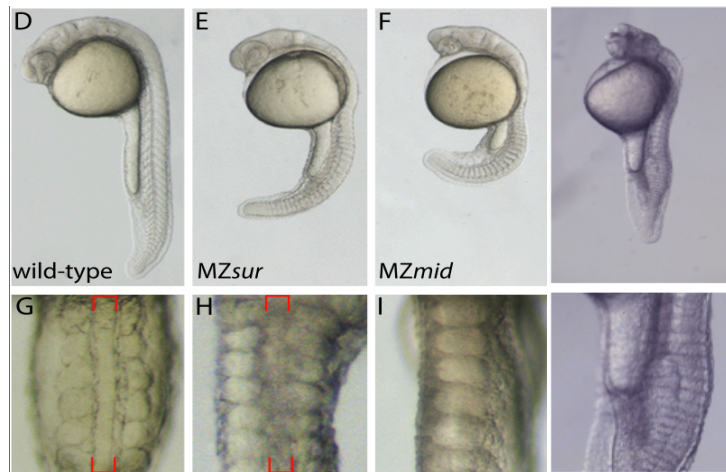
B) Front head view: closely set eyes, partial or complete cyclopia is present in all MZ mutants.

C) Side trunk view: poorly developed somites and absence of notochord are visible.

Among several clutches observed (~300 embryos in total), no embryo without any phenotypic abnormalities was observed, and none survived past 4 days. The defects that we observed were not significantly more severe than the ones described in the *midway* mutants [Fig.2.6]. Considering the fact that both the

previously generated mutant *midway* and the complete FoxH1 *E1* mutant failed to exhibit a pre-gastrulation phenotype, we can conclude that the morpholino phenotype is likely not a consequence of FoxH1 missing, but rather a side-effect of the injection of this morpholino.

A. wild-type || *MZsur* || *MZmid* || *MZ E1* mutant



B. wild-type *MZE1* mutant wild-type *MZE1* mutant

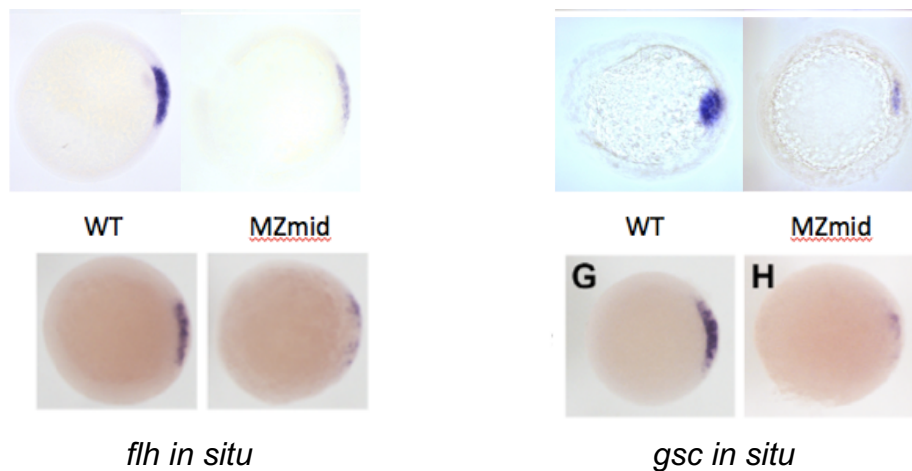


Figure 2.6

Side-by-side comparison of the MZ midway mutant and MZE1 mutant

A) Whole-body (above) and trunk (below) phenotype of wild-type, *MZsur*, *MZmid* (from (Slagle et al., 2011)) and *MZE1* embryos.

B) *in situ* hybridization pictures of the expression of two Nodal-regulated genes, *flh* and *gsc*, at shield stage. Wild-type and *MZE1* mutant embryos above, wild-type and *MZmid* embryos below (from (Slagle et al., 2011)).

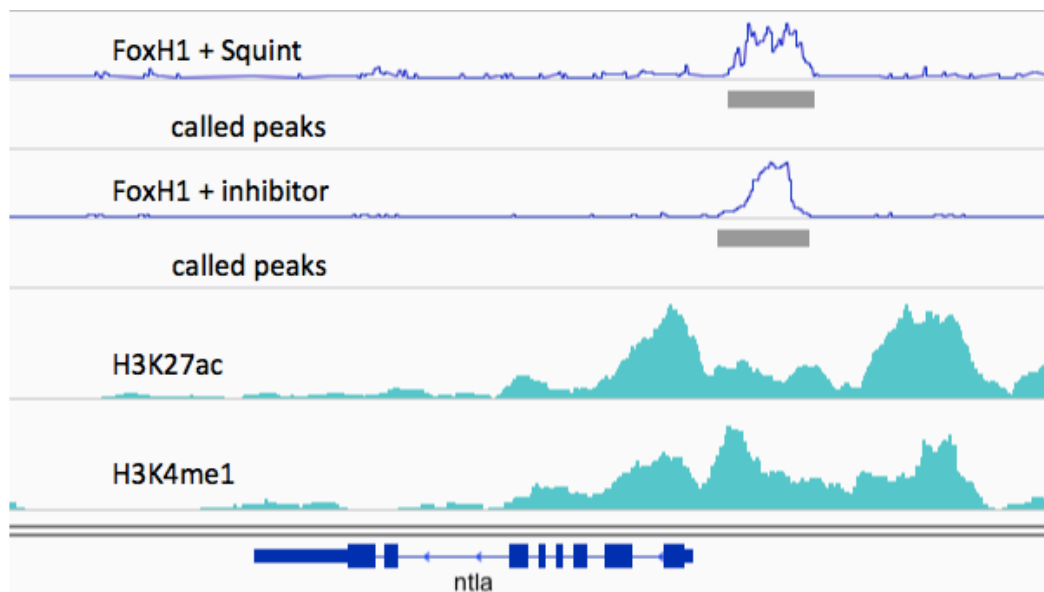
Identification of the direct targets of FoxH1 in zebrafish

Although FoxH1 has long been known as a transcription factor crucial for early development in all vertebrates, only a limited number of direct targets of this factor were identified in zebrafish (Sirotkin et al., 2000; Slagle et al., 2011). Several papers proposed that FoxH1 in other organisms might have a function in development other than the recruitment of Smad2 to Nodal target genes (Chiu et al., 2014; Silvestri et al., 2008). Only a few of those, however, have been found in zebrafish so far (Pei et al., 2007; Slagle et al., 2011).

In order to identify many or all FoxH1 targets in zebrafish embryo, I sequenced mRNA from wild type and *MZEI* mutant embryos at dome stage (4hpf) using deep sequencing. After analyzing the data, I obtained a list of 1660 unique genes with significantly different expression levels ($q < 0.05$) between these two groups. Applying the cutoff of 2-fold difference in either direction (increase or decrease in expression), 92 genes were found to be upregulated in *MZEI* mutant embryos, while 457 genes were downregulated.

To find out which of these genes are directly regulated by FoxH1, I performed ChIP-Seq of FoxH1 transcription factor using dome stage embryos. I performed chromatin immunoprecipitation (ChIP) with an anti-flag antibody in embryos expressing FoxH1-flag construct. Nodal signaling in the analyzed embryos was either blocked by drug addition (DR), overactivated by the injection of high levels of *squint* mRNA (SQ), or unperturbed (WT). The validity of the data is supported by the fact that FoxH1 binding peaks are enriched in the proximity of known Nodal target genes, and overlap with enhancer-specific chromatin marks (H3K4me1, H3K27ac (data from (Bogdanovic et al., 2012)) [Fig.2.7].

Gene name: *ntla*



Gene name: *lft2*

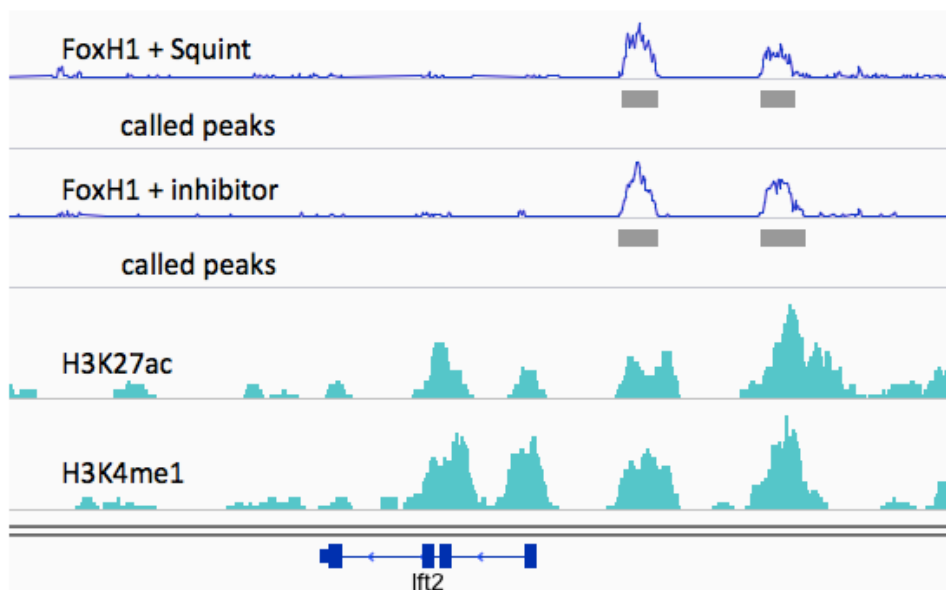


Figure 2.7 (continued on the next page)

Figure 2.7 (continued)

Gene name: *lhx1a*

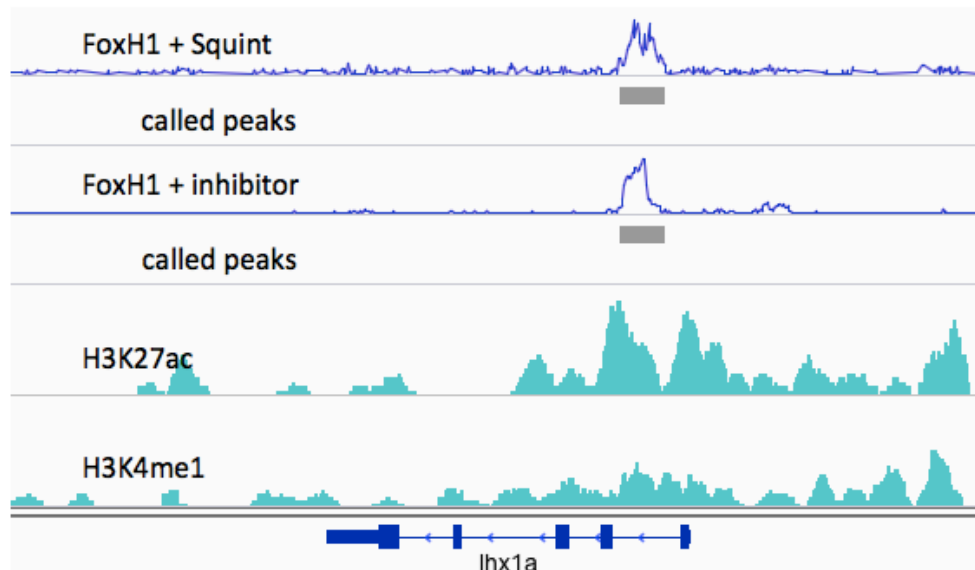


Figure 2.7

ChIP tracks for FoxH1 transcription factor binding and enhancer chromatin marks

For each gene, top to bottom, the following tracks are shown:

FoxH1 factor binding in embryos coinjected with *squint* mRNA; results of peak calling on this track; FoxH1 factor binding in embryos treated with Nodal drug inhibitor; results of peak calling on this track; H3K27ac and H3K4me1 ChIP tracks at dome stage (from (Bogdanovic et al., 2012)).

Motif discovery tools identify FoxH1 consensus binding sequence in the FoxH1 peaks, as well as those of Oct-family proteins and Smad2/3 binding sequences [Fig.2.8]. Of note, 15.4% of FoxH1 binding peaks also contained the common binding motif for Zic1/2/3 transcription factors [Fig.2.8]. Zic family of zinc-finger proteins has been shown to play an important role in animal development. In pluripotent cells Zic3 is required for maintenance of stem cells pluripotency upon binding to promoters of genes associated with cell pluripotency (Nanog, Sox2, Oct4, etc.) as well as cell cycle, proliferation, oncogenesis and early embryogenesis (Lim et al., 2010, 2007). In contrast, during gastrulation and neurulation Zic3 acts by binding the enhancers associated with control of gene transcription in the Nodal and Wnt

signaling pathways (Winata et al., 2013). Zic family proteins also often bind to adjacent sites in the context of multi-transcription factor complexes associated with regulatory elements.

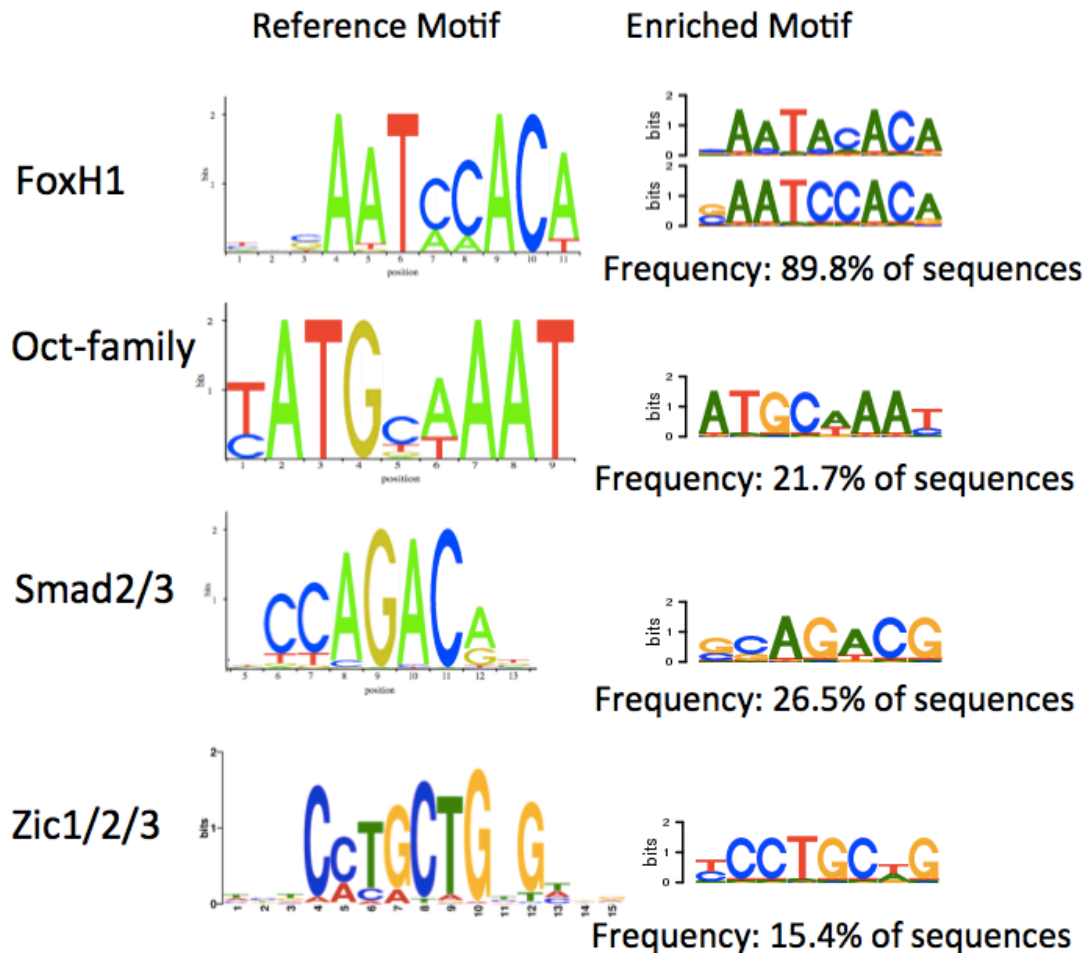


Figure 2.8

FoxH1 peaks de-novo motif analysis

Transcription factor binding motif analysis identified consensus binding sequences for FoxH1, Smad2/3, Zic and Oct-family members.

ChIP-Seq data from dome stage drug-treated embryos confirms that in most cases FoxH1 binding to the regulatory elements does not require Nodal signaling. The fact that FoxH1 is capable of binding to DNA early on in development,

and before any exposure to the Nodal signal, and can later recruit the activating R-Smad complex, indicates that FoxH1 can be considered a pioneer factor.

We were interested in determining whether there is a correlation between FoxH1 binding and the observed threshold of the nearest gene activation. I used MACS2 software to identify the number and position of FoxH1 peaks near genes that are low and high threshold. I did not find a significant difference in number and position of such peaks between these groups of genes. This suggests that FoxH1 doesn't play an important role in the determination of activation threshold of Nodal target genes. It is likely that other transcription factors play a role in this regulation.

FoxH1 functions in the regulation of genes that are not Nodal targets

Out of the 549 genes that changed their expression in *MZE1* mutant embryos, 84 genes had FoxH1 peak within 20kb of their transcription start site (TSS), suggesting that these are direct FoxH1 targets [Table 2.1]. Of these, 13 were upregulated and 71 downregulated in *MZE1* mutants. 25 of these 84 genes also had a Smad2 peak within 20kb of the TSS, among them well known Nodal target genes, such as *gsc*, *lft1*, *foxa2*. In addition, 11 of these 84 genes were identified as Nodal-regulated by previous research based on expression studies (Bennett et al., 2007). This adds up to 30 genes, where the main role of FoxH1 is most probably the recruitment of activated Smad2. Gene ontology analysis of this group showed that the following groups of genes were overrepresented, among others: axis elongation (GO:0003401); regionalization (GO:0003002); pattern specification process (GO:0007389); cell differentiation (GO:0030154). These groups were not overrepresented in the analysis of FoxH1 regulated genes that are not Nodal target genes. In contrast, the following classes of genes were overrepresented: cell morphogenesis involved in neuron

differentiation (GO:0048667); axon development (GO:0061564); cell migration (GO:0016477); neuron projection development (GO:0031175); localization of cell (GO:0051674); cell motility (GO:0048870).

Table 2.1
FoxH1 direct target genes, based on RNA-Seq and ChIP-Seq analysis

Gene name	fold change log2	Gene name	fold change log2	Gene name	fold change log2
angptl4	-4.01477	alcamb	2.15064	serinc5	2.88414
her3	-3.72952	lemd3	2.1873	fzd7b	2.91004
krt91	-2.75941	sec61a1l	2.19379	tuba1a	2.9283
lim2.3	-2.61858	fgfr1b	2.19773	mfsd11	2.9562
her7	-2.44526	sdc4	2.19815	amfr	2.98414
klf17	-2.22545	lpgat1	2.24156	serpina10a	3.0143
nfil3	-2.19859	ptger4b	2.29186	re1t	3.04513
znf143b	-2.19601	aoc2	2.29757	erlin1	3.08424
meig1	-2.14454	itm2cb	2.30848	igf2a	3.16639
h2afx	-2.13372	cxc4b	2.31706	tmbim1	3.21415
lama4	-2.11796	p4ha2	2.32752	atp1b1a	3.22063
her6	-2.10744	bbc3	2.33893	slc16a1b	3.37816
krt18	-2.07604	acvr2aa	2.3458	cdkn1a	3.43378
gsc	1.91994	atp1a1a.1	2.36129	pkd2	3.50687
aplnra	1.92861	itgb5	2.38941	itga5	3.52659
kif1b	1.94021	tspan7b	2.4009	sfrp1a	3.55539
susd6	1.95375	bcap29	2.41366	gpc4	3.66187
hmgxb4a	2.0121	slc35e3	2.42683	abca1a	3.69615
gde1	2.01453	tmem2	2.42891	cntn2	3.82175
spint2	2.01547	acvr1ba	2.42922	adam10a	3.83637
wee2	2.07031	kazald2	2.47959	notch3	3.90263
golim4b	2.0754	ctsc	2.53773	aplnrb	4.0503
bri3bp	2.08561	bmp7a	2.54617	foxa2	4.09775
tmem9b	2.10248	alpl	2.56549	lamc1	4.37879
slc39a6	2.10968	cx43.4	2.57884	clstn1	4.73137
dhhs9	2.10976	gylt11b	2.61633	notum1a	5.51099
tph1b	2.12944	glipr2l	2.70213	ptk7a	5.54227
cmtm7	2.13001	ptprfb	2.77218	lft1	5.73446

These results are in agreement with the results observed in other organisms and in FoxH1 morphants in zebrafish. One study in mouse embryos revealed that FoxH1 initiates a transcriptional regulatory network within the developing anterior neuroectoderm and plays a role in initiating retinoic acid signaling in the forebrain (Silvestri et al., 2008). Another recent study in *Xenopus* embryos identified 72 FoxH1 targets that lack Smad2/3 binding and are insensitive to Nodal signaling inhibition. 27 of these genes are upregulated in the absence of FoxH1, suggesting that FoxH1 functions as a dual-acting TF to activate or repress target genes in a Nodal-independent manner (Chiu et al., 2014).

In zebrafish, apart from disrupting epiboly, FoxH1 MO treatment disrupted convergence and internalization movements and disrupted the expression of five keratin genes – *cyt1*, *cyt2*, *krt4*, *krt8* and *krt18* - that are normally transcribed in the embryo's enveloping layer. Our results found two keratin genes to be disrupted: *krt18* and *krt91*. However, in contrast to the previous results, we did not observe severe abnormalities in cell movement during epiboly. In fact, the morphology was normal until shield stage. At 24hpf, the development of brain, eyes and somites was disrupted in *MZE1* mutants, but not as dramatically as in morphants (Pei et al., 2007). Similar to the results in mice, FoxH1 was shown by us to regulate a number of genes that are important in the development of the nervous system. As in *Xenopus* embryos, FoxH1 was shown by our study to downregulate a number of genes. 8 out of the 13 downregulated genes are transcription factors, functioning in the development of heart, brain, and blood circulation. 3 more encoded structural proteins in epidermal cells, blood vessels, and lens.

DISCUSSION

FoxH1 has been identified as one of the key factors in mesendoderm development. Although mutations of this transcription factor do not abrogate Nodal signaling entirely, they lead to severe loss of mesendodermal tissues (Hoodless et al., 2001; Yamamoto et al., 2001). Two different mutants and a morphant have been generated for FoxH1 in zebrafish, but it was still unclear whether complete loss of its function had been achieved. Using Cas9 technology, we have been able to generate alleles with a frame-shift mutation early in the transcript for in depth analysis of this transcription factor.

We identified a number of FoxH1 target genes that are not Nodal-regulated, while we didn't detect any phenotypic defects in the FoxH1 mutant that were drastically different from those in *MZoepe* mutants. It must be noted, however, that many of the Nodal-independent FoxH1 target genes are involved in the development of the nervous system. Since early defects in Nodal signaling transduction lead to abnormalities in head and brain development, these could easily mask the later Nodal-independent role of FoxH1. Our analysis of FoxH1 target genes also identified two keratin genes: *krt18* and *krt91*, similarly to the findings in FoxH1 morphant (Pei et al., 2007). Contrary to their results, however, we did not observe a disruption in epiboly, convergence and internalization movements. It can be explained by the fact that morpholino decreased the expression of five keratin genes – *cyt1*, *cyt2*, *krt4*, *krt8* and *krt18* – unlike the two in FoxH1 mutant. We hypothesize that morpholino had off-target effects on keratin genes that we did not find to be disrupted in the mutant. While other keratin genes can compensate for the loss of *krt18* and *krt91* in FoxH1 mutants, they are unable to do so when FoxH1 morpholino was injected.

In addition to the 71 genes that had a FoxH1 binding site near their TSS and were downregulated in FoxH1 mutants, we found 13 that were upregulated, suggesting that FoxH1 can also act as a repressor. 8 out of these repressed genes are transcription factors, functioning in the development of heart, brain, and blood circulation, and 3 are structural proteins in epidermal cells, blood vessels, and lens. This is consistent with previous observations that overexpression of FoxH1 has a negative effect on vascular formation in zebrafish embryos (Choi et al., 2007).

The mechanism of FoxH1 repressive activity has to be studied further. Many transcription factors in fork head family of proteins can mediate gene repression through recruitment of TLE/Groucho family transcriptional corepressors (Yaklichkin et al., 2007) and FoxH1 has a conserved Engrailed homology-1 (EH1) motif. Studies in mouse, however, showed that disrupting FoxH1–Groucho interaction through mutations in this domain do not lead to significant gene expression changes (Halstead and Wright, 2015). While it is possible that recruitment of transcriptional repressors can happen without the EH1 domain, it is also possible that FoxH1 competes for binding with activating factors, and represses gene expression in this manner.

To regulate Nodal-independent genes, FoxH1 should be capable of binding to DNA in the absence of Nodal signaling. This is indeed what we observed in our ChIP experiments. The majority of FoxH1 binding peaks were present with or without Nodal signaling. One caveat of our experiments is that we didn't use an antibody against the endogenous FoxH1 protein. Testing showed that the levels of *FoxH1-flag* mRNA we injected didn't cause any developmental abnormalities, suggesting that we didn't flood the embryo with FoxH1, which could have led to improper binding. However, using an antibody against endogenous FoxH1 or using

transgenic zebrafish that express a *FoxH1-flag* construct under a native promoter would be ideal for further experiments with this transcription factor. This setup could be used to determine the earliest point at which FoxH1 binds to its target loci, since maternally provided FoxH1 should be present in the embryos long before zygotic genome activation.

The fact that FoxH1 binds to enhancers without Nodal signaling and later recruits Smad2 to these loci suggests that FoxH1 plays the role of a pioneer factor. Previously identified pioneer factors often play roles in modifying/remodeling histones and changing/maintaining DNA methylation status (Heinz et al., 2010; Xu et al., 2009, 2007). ChIP studies in FoxH1 mutants can show whether it plays similar roles in regulating chromatin landscape. DNA methylation and nucleosome density profile, the status of H3K4me1, H3K4me2, and H3K9me3 marks can be compared between wild-type embryos and FoxH1 mutants.

Our study identified a number of novel FoxH1 target genes, some of them activated and some repressed by this transcription factor. We also found 54 FoxH1 target genes that are not Nodal-regulated. Study of FoxH1 binding sites helped us identify other transcription factors that it potentially interacts with, and gave us data on which Smad2 peaks are associated with FoxH1 binding, which is important for better understanding of Nodal signaling.

MATERIALS AND METHODS

FoxH1 mutant generation:

Target sites that matched the sequence GG-N19-GG, GA-N19-GG, or AG- N19-GG were selected in the first, second and third exon of the FoxH1 gene, and checked for uniqueness in the genome. To generate templates for sgRNA transcription, gene-specific oligonucleotides containing the T7 promoter sequence, the 20 base target site without the PAM, and a complementary region were annealed to a constant oligonucleotide encoding the reverse-complement of the tracrRNA tail. sgRNA templates were purified using Qiaquick columns (Qiagen) and transcribed using Megascript kit (Ambion). All sgRNAs were then DNase treated and precipitated with ammonium acetate/ethanol. Cas9 mRNA was transcribed from linearized template DNA using mMachine SP6 kit (Ambion), DNase treated, and precipitated with lithium chloride. Zebrafish TLAB strain zygotes were collected and injected through the chorion with a mix of 25 pg sgRNA, 300 pg Cas9 mRNA, and phenol red dye in a single mix. The resulting F0 generation zebrafish were grown to adulthood, and their progeny from a cross to WT fish were sorted using MiSeq to identify frame-shifting indel mutations. F1 embryos containing the required mutations were grown to adulthood and crossed to each other to obtain homozygous fish. To rescue zygotic homozygous FoxH1 embryos, 10pg of FoxH1 mRNA was injected through the chorion at one-cell stage. FoxH1 mRNA was obtained from full-length zebrafish FoxH1 cDNA cloned into pcs2+ plasmid and transcribed using mMachine SP6 kit (Ambion).

Chromatin Immunoprecipitation (ChIP)

Embryos for ChIP were collected at dome stage after 5 pg squint mRNA injection or after treatment with the Nodal signaling inhibitor SB505124 (Sigma S4696) at 20 μ M final. For FoxH1 ChIP, embryos were injected with 5 pg of FoxH1-flag mRNA at 1-cell stage, and anti-flag antibody was used for the pull down. For each ChIP, 800 embryos were collected and fixed in 1.85% formaldehyde for 15 min at 20°C. Formaldehyde was quenched by adding glycine to a final concentration of 0.125 M. Embryos were rinsed three times in ice-cold PBS, and resuspended in cell lysis buffer (10 mM Tris-HCl pH7.5/10 mM NaCl/0.5% NP40) and lysed for 15 min on ice. Nuclei were collected by centrifugation, resuspended in nuclei lysis buffer (50mMTris-HCl pH 7.5/10mMEDTA/1% SDS) and lysed for 10 min on ice. Samples were diluted three times in IP dilution buffer (16.7 mM Tris-HCl pH 7.5/167 mM NaCl/1.2 mM EDTA/ 0.01% SDS) and sonicated to obtain fragments of \sim 500 bp. Triton X-100 was added to a final concentration of 0.75% and the lysate was incubated overnight while rotating at 4°C with 25 μ lof protein G magnetic Dynabeads (Invitrogen) pre-bound to an excess amount of antibody. Antibody used was anti-FLAG M1 (Sigma F3165). Bound complexes were washed six times with RIPA (50 mM HEPES pH7.6/1 mM EDTA/0.7% DOC/1% Igepal/0.5 M LiCl) and TBS and then eluted from the beads with elution buffer (50 mM NaHCO₃/1% SDS). Crosslinks were reversed overnight at 65°C and DNA purified by the QIAquick PCR purification kit (Qiagen).

ChIP-qPCR

ChIP was done as described above. Embryos were injected with 5 pg of FoxH1-flag mRNA at 1-cell stage, and collected at 1K-cell stage. The eluted DNA was analyzed

using SYBR[®] Green I (BioRad) and DNA Engine Opticon System (MJ Research) with technical duplicates. ChIP enrichment was normalized to input, and plotted as fold enrichment over input.

ChIP-Seq

ChIP was performed as described above. Embryos were untreated, incubated at 1-cell stage with 10mM Nodal inhibitor drug SB-505124 (DaCosta Byfield et al., 2004) (S4696 SIGMA), or injected at 1-cell stage with 5pg of *squint* mRNA. Libraries were prepared according to the Illumina sequencing library preparation protocol and sequenced on an Illumina HiSeq 2000. Reads were aligned to *Danio rerio* genome Zv9 assembly using. ChIP-seq reads were mapped using Bowtie (v.0.12.9) (Langmead, 2010); peaks were called using MACS (v.2.0.10) (Zhang et al., 2008). De novo motif analyses used XXMOTIF (Hartmann et al., 2013). JASPAR public database was used to find candidate proteins. GO analyses were performed using PANTHER (protein annotation through evolutionary relationship) classification system (Mi et al., 2013).

RNA-Seq

Approximately 100 embryos at dome stage were dechorionated, and whole RNA was extracted from them using TRIZOL, according to the manufacturer's protocol. mRNA was isolated using poly-A tail isolation method, libraries were prepared according to the Illumina sequencing library preparation protocol, and sequenced on an Illumina HiSeq 2000. Differential gene expression analysis of RNA-seq data was performed using TopHat (v. 1.3.3) (Trapnell et al., 2009) and Cuffdiff (v. 1.3.0) (Trapnell et al.,

2012). GO analyses were performed using PANTHER (protein annotation through evolutionary relationship) classification system (Mi et al., 2013).

RT-qPCR

For gene expression analysis, total RNA was isolated from approximately 20 zebrafish embryos using QIAgene RNeasy kit. cDNA was made from total RNA using BioRad iScript kit, and analyzed using SYBR[®] Green I (BioRad) and DNA Engine Opticon System (MJ Research). A housekeeping gene β -actin was used for normalization across samples.

In situ hybridization

In situ hybridization on whole mount embryos was carried out using standard methods. Embryos were fixed in 4% formaldehyde overnight, and then dehydrated, rehydrated, and stained. Following staining, embryos were cleared in benzyl benzoate/benzyl alcohol (2:1 vol/vol) and imaged.

REFERENCES

- Bennett, J.T., Joubin, K., Cheng, S., Aanstad, P., Herwig, R., Clark, M., Lehrach, H., Schier, A.F., 2007. Nodal signaling activates differentiation genes during zebrafish gastrulation. *Dev. Biol.* 304, 525–540. doi:10.1016/j.ydbio.2007.01.012
- Bogdanovic, O., Fernandez-Minan, A., Tena, J.J., De La Calle-Mustienes, E., Hidalgo, C., Van Kruysbergen, I., Van Heeringen, S.J., Veenstra, G.J.C., Gomez-Skarmeta, J.L., 2012. Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Res.* 22, 2043–2053. doi:10.1101/gr.134833.111
- Chen, X., Rubock, M.J., Whitman, M., 1996. A transcriptional partner for MAD proteins in TGF-beta signalling. *Nature*. doi:10.1038/383691a0
- Chiu, W.T., Charney Le, R., Blitz, I.L., Fish, M.B., Li, Y., Biesinger, J., Xie, X., Cho, K.W.Y., 2014. Genome-wide view of TGFβ/Foxh1 regulation of the early mesendoderm program. *Development* 141, 4537–47. doi:10.1242/dev.107227
- Choi, J., Dong, L., Ahn, J., Dao, D., Hammerschmidt, M., Chen, J.N., 2007. FoxH1 negatively modulates flk1 gene expression and vascular formation in zebrafish. *Dev. Biol.* 304, 735–744. doi:10.1016/j.ydbio.2007.01.023
- DaCosta Byfield, S., Major, C., Laping, N.J., Roberts, A.B., 2004. SB-505124 is a selective inhibitor of transforming growth factor-beta type I receptors ALK4, ALK5, and ALK7. *Mol. Pharmacol.* 65, 744–752. doi:10.1124/mol.65.3.744
- Dubrulle, J., Jordan, B.M., Akhmetova, L., Farrell, J.A., Kim, S.-H., Solnica-Krezel, L., Schier, A.F., 2015. Response to Nodal morphogen gradient is determined by the kinetics of target gene induction. *Elife* 4. doi:10.7554/eLife.05042
- Gagnon, J.A., Valen, E., Thyme, S.B., Huang, P., Ahkmetova, L., Pauli, A., Montague, T.G., Zimmerman, S., Richter, C., Schier, A.F., 2014. Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS One* 9. doi:10.1371/journal.pone.0098186
- Halstead, A.M., Wright, C.V.E., 2015. Disrupting Foxh1-Groucho interaction reveals robustness of nodal-based embryonic patterning. *Mech. Dev.* 136, 155–165. doi:10.1016/j.mod.2014.12.002
- Hartmann, H., Guthohrlein, E.W., Siebert, M., Luehr, S., Soding, J., 2013. P-value-based regulatory motif discovery using positional weight matrices. *Genome Res.* 23, 181–194. doi:10.1101/gr.139881.112
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., Glass, C.K., 2010. Simple Combinations of Lineage-

- Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* 38, 576–589. doi:10.1016/j.molcel.2010.05.004
- Hoodless, P.A., Pye, M., Chazaud, C., Labbé, E., Attisano, L., Rossant, J., Wrana, J.L., 2001. FoxH1 (Fast) functions to specify the anterior primitive streak in the mouse. *Genes Dev.* 15, 1257–1271. doi:10.1101/gad.881501
- Kim, S.W., Yoon, S.J., Chuong, E., Oyolu, C., Wills, A.E., Gupta, R., Baker, J., 2011. Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. *Dev. Biol.* 357, 492–504. doi:10.1016/j.ydbio.2011.06.009
- Kunwar, P.S., Zimmerman, S., Bennett, J.T., Chen, Y., Whitman, M., Schier, A.F., 2003. Mixer/Bon and FoxH1/Sur have overlapping and divergent roles in Nodal signaling and mesendoderm induction. *Development* 130, 5589–5599. doi:10.1242/dev.00803
- Langmead, B., 2010. Aligning short sequencing reads with Bowtie. *Curr. Protoc. Bioinforma.* doi:10.1002/0471250953.bi1107s32
- Lim, L.S., Hong, F.H., Kunarso, G., Stanton, L.W., 2010. The pluripotency regulator Zic3 is a direct activator of the Nanog promoter in ESCs. *Stem Cells* 28, 1961–1969. doi:10.1002/stem.527
- Lim, L.S., Loh, Y.-H., Zhang, W., Li, Y., Chen, X., Wang, Y., Bakre, M., Ng, H.-H., Stanton, L.W., 2007. Zic3 is required for maintenance of pluripotency in embryonic stem cells. *Mol. Biol. Cell* 18, 1348–58. doi:10.1091/mbc.E06-07-0624
- Mi, H., Muruganujan, A., Casagrande, J.T., Thomas, P.D., 2013. Large-scale gene function analysis with the PANTHER classification system. *Nat. Protoc.* 8, 1551–66. doi:10.1038/nprot.2013.092
- Pei, W., Noushmehr, H., Costa, J., Ouspenskaia, M. V., Elkahloun, A.G., Feldman, B., 2007. An early requirement for maternal FoxH1 during zebrafish gastrulation. *Dev. Biol.* 310, 10–22. doi:10.1016/j.ydbio.2007.07.011
- Silvestri, C., Narimatsu, M., von Both, I., Liu, Y., Tan, N.B.J., Izzi, L., McCaffery, P., Wrana, J.L., Attisano, L., 2008. Genome-wide identification of Smad/Foxh1 targets reveals a role for Foxh1 in retinoic acid regulation and forebrain development. *Dev. Cell* 14, 411–423. doi:10.1016/j.devcel.2008.01.004
- Sirotkin, H.I., Gates, M.A., Kelly, P.D., Schier, A.F., Talbot, W.S., 2000. fast1 is required for the development of dorsal axial structures in zebrafish. *Curr. Biol.* 10, 1051–1054. doi:10.1016/S0960-9822(00)00679-5
- Slagle, C.E., Aoki, T., Burdine, R.D., 2011. Nodal-dependent mesendoderm specification requires the combinatorial activities of FoxH1 and eomesodermin.

PLoS Genet. 7. doi:10.1371/journal.pgen.1002072

- Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111. doi:10.1093/bioinformatics/btp120
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., Pachter, L., 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–78. doi:10.1038/nprot.2012.016
- Winata, C.L., Kondrychyn, I., Kumar, V., Srinivasan, K.G., Orlov, Y., Ravishankar, A., Prabhakar, S., Stanton, L.W., Korzh, V., Mathavan, S., 2013. Genome Wide Analysis Reveals Zic3 Interaction with Distal Regulatory Elements of Stage Specific Developmental Genes in Zebrafish. *PLoS Genet.* 9. doi:10.1371/journal.pgen.1003852
- Xu, J., Pope, S.D., Jazirehi, A.R., Attema, J.L., Papathanasiou, P., Watts, J. a, Zaret, K.S., Weissman, I.L., Smale, S.T., 2007. Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 104, 12377–12382. doi:10.1073/pnas.0704579104
- Xu, J., Watts, J.A., Pope, S.D., Gadue, P., Kamps, M., Plath, K., Zaret, K.S., Smale, S.T., 2009. Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells. *Genes Dev.* 23, 2824–2838. doi:10.1101/gad.1861209
- Yaklichkin, S., Vekker, A., Stayrook, S., Lewis, M., Kessler, D.S., 2007. Prevalence of the EH1 Groucho interaction motif in the metazoan Fox family of transcriptional regulators. *BMC Genomics* 8, 201. doi:1471-2164-8-201 [pii]r10.1186/1471-2164-8-201
- Yamamoto, M., Meno, C., Sakai, Y., Shiratori, H., Mochida, K., Ikawa, Y., Saijoh, Y., Hamada, H., 2001. The transcription factor FoxH1 (FAST) mediates Nodal signaling during anterior-posterior patterning and node formation in the mouse. *Genes Dev.* 15, 1242–1256. doi:10.1101/gad.883901
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., Liu, X.S., 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137. doi:10.1186/gb-2008-9-9-r137

CHAPTER III

Identification of Smad2 binding sites and regulatory elements

ABSTRACT

Identifying regulatory elements that recruit Smad2 is crucial for our understanding of Nodal signaling pathway. In this Chapter I describe experiments performed towards this goal. I carried out Smad2 ChIP-Seq in embryos overexpressing Nodal signal Squint, thus detecting loci that bind Smad2 after exposure to high Nodal levels. I found 1041 such peaks, enriched for the binding motifs of Smad2/3, FoxH1, EomesA, Oct and Homeobox family proteins, as well as a negative regulator Osr1. I tested DNA elements of approximately 500bp centered on a Smad2 binding peak for their gene regulatory potential, and discovered that they are sufficient to drive gene expression in a Nodal-dependent manner. I also identified 26 previously unpublished Nodal target genes, and 55 genes that are bound by Smad2 upon exposure to high, but not low levels of Nodal signaling.

PREFACE

I designed, performed, and interpreted all experiments. Andrea Pauli generated RNA-Seq data in Squint-injected and MZ*oep* zebrafish. The Gomez-Skarmeta lab generously provided ZED vector. Smad2 and FoxH1 ChIP-Seq data has been previously published in (Dubrulle et al., 2015). Alexander F. Schier helped conceive, design, and support the project.

INTRODUCTION

The recruitment of phosphorylated Smad2 to DNA is considered to be the final event necessary for the activation of Nodal-regulated genes, an on-off switch for gene transcription initiation. It has therefore been assumed (but not proven) that binding conditions of Smad2 transcription factor are different between genes with higher vs. lower threshold of activation. If this assumption is true, other factors should be responsible for determining the conditions under which Smad2 will be recruited to the locus, since Smad2 is incapable of binding to DNA by itself. These determining factors could be the sequence-specific co-transcription factors, such as FoxH1 and Mixer, or the chromatin landscape of the locus, including nucleosome occupancy.

Chromatin immunoprecipitation (ChIP) has been used in many systems and for many DNA-binding proteins to understand the binding dynamics of a factor, as well as to identify the direct target genes that they regulate. Even though Smad2 is one of the most important transcription factors in early vertebrate development, few studies addressed the binding patterns of this factor, partly because of the relative difficulty of the pull down, stemming from the fact that Smad2 doesn't bind to DNA directly. Six papers employing Smad2 ChIP have so far been published. Three of these performed ChIP in human or mouse cell culture (Brown et al., 2011; Fei et al., 2010; Kim et al., 2011), and the most recent one in *Xenopus* embryo (Chiu et al., 2014).

Two previous studies (Liu et al., 2011; Nelson et al., 2014) described Smad2 ChIP results in zebrafish embryos. The first of those two studies used untreated early gastrula (high-sphere stage) embryos for their Smad2 pulldown, and anti-Smad2/3 (3102, Cell Signaling Technology) antibody. For the analysis of the pulled-down DNA, they used The Zebrafish Promoter Genome ChIP-on-chip

Microarray Set (Agilent Technologies). This oligonucleotide-based promoter array set covers 9kb upstream of and 3kb downstream of the transcription start sites of 11,512 zebrafish genes. Since many developmental enhancers can be as far as several hundred kilobases from the transcription start site of gene that they regulate, this study was unable to discover many Smad2 binding sites. The researchers found 556 Smad2 peaks, and only 14 of those were close to well-known Nodal regulated genes.

The second study in 2014 used 5000 wild type zebrafish embryos at early gastrula stage (high stage) and Smad2 monoclonal antibody (clone 31H15L4; Thermo Fisher Scientific) for their pulldown. High-throughput DNA sequencing was used to analyze the acquired DNA. This group was successful in obtaining good genome coverage, and identifying a large number of Smad2 binding sites in proximity to known Nodal target genes. Using RNA-Seq, they also identified a number of new potential Nodal signaling targets. To identify Smad2 peaks with high level of certainty, it is very useful to know where its co-factors (FoxH1, Mixer, eomesA) bind in the genome. The Wardle lab performed eomesA ChIP-Seq as well as Smad2 ChIP-Seq (Nelson et al., 2014), but no FoxH1 ChIP-Seq data in zebrafish embryos has been previously published. The dataset that I acquired (see Chapter II) is therefore important for the deeper understanding of how Smad2 binding in response to Nodal signaling is controlled during vertebrate embryonic development.

The caveat of previously published Smad2 binding data in zebrafish is that both groups used untreated, wild type embryos for their work. Wild-type embryos are a mixture of cells with a very high number of cells not experiencing any Nodal signaling, a few that are exposed to low levels of it, and very few exposed to high Nodal levels. We can therefore suppose that many binding sites that are only occupied when a cell is exposed to very high signal levels would be overlooked with this setup.

Another concern is that the stage that both groups selected (high-sphere and high stage, respectively) happens very early in the development, at the very onset of Nodal signaling. Thus, it is possible that enhancers that are occupied by Smad2 only in the presence of high levels of Nodal and/or after long exposure to the signal would not be detected. I carried out Smad2 pull-down experiments in embryos injected with *squint* mRNA to find out the regulatory elements that are bound at high Nodal signaling levels, and analyzed their responsiveness to Nodal signaling.

RESULTS

Identification of Smad2-bound loci and co-bound transcription factor candidates

To activate Nodal signaling uniformly, I used embryos that were injected with 5pg of *squint* mRNA at one-cell stage. For comparison, I also performed ChIP-Seq in embryos treated with a Nodal signaling inhibitor. I also chose to study a later developmental stage (dome) than the ones used in the two previously published papers. The later developmental stage and the fact that the extracellular space in the embryos is flooded with Nodal signaling molecules from the moment *squint* mRNA starts being translated, means that genes requiring long exposure to the signal will be activated by the time the embryos are collected.

Legend

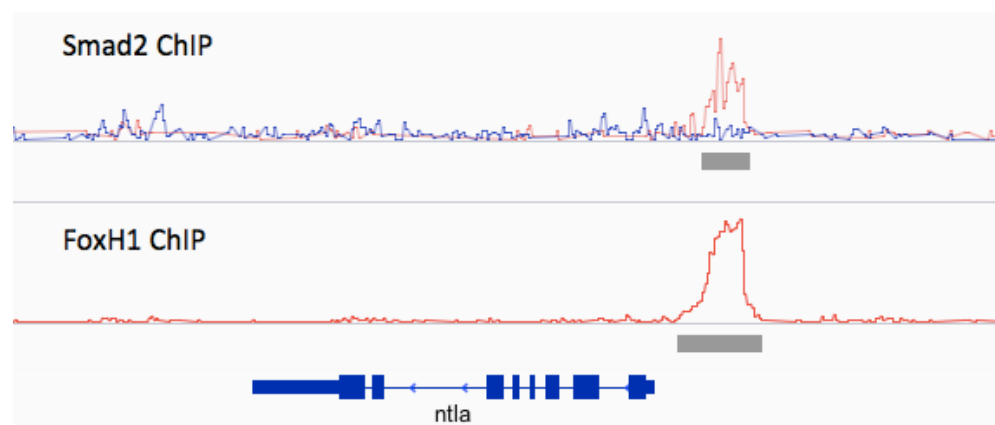
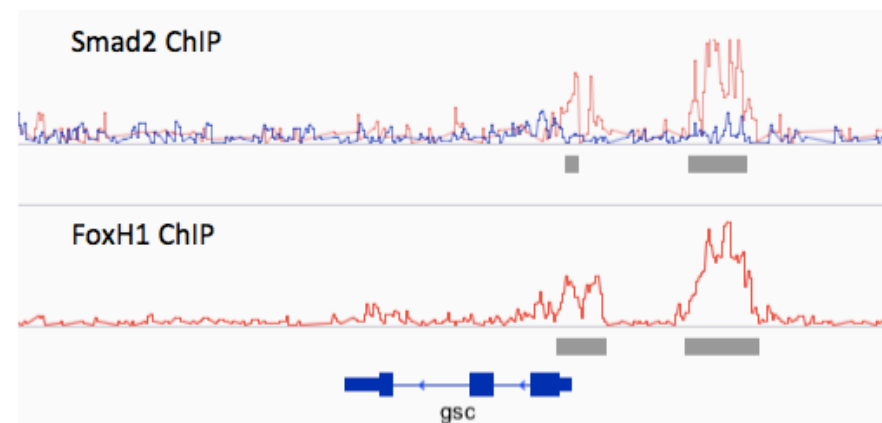
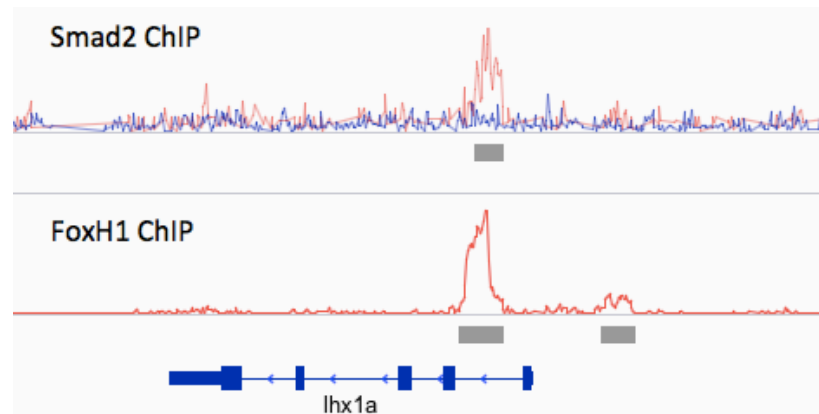


Figure 3.1 (continued on the next page)

Figure 3.1 (continued)

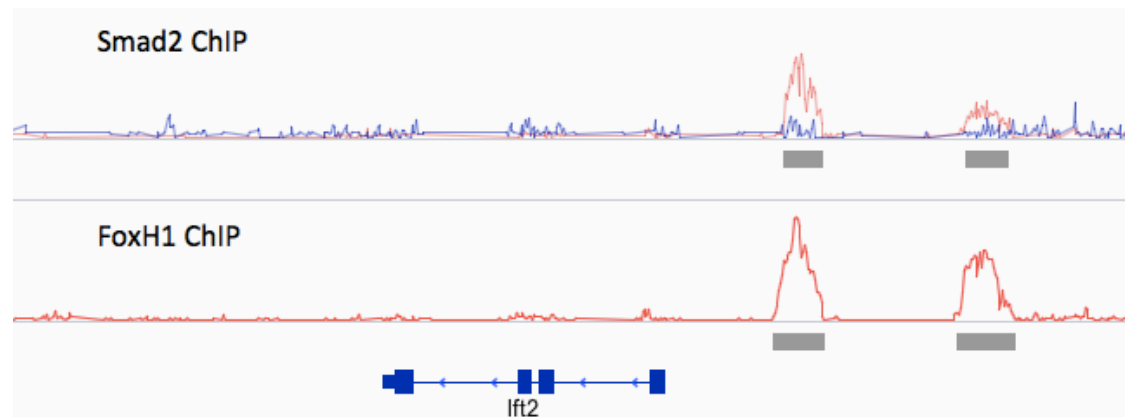


Figure 3.1

Smad2 transcription factor binds to regulatory elements in a Nodal-dependent manner

For each of the shown Nodal target genes (*lhx1a*, *gsc*, *ntla*, *lft2*) Smad2 and FoxH1 ChIP-Seq tracks are shown.

Tracks from top to bottom: overlay of Smad2 ChIP-Seq track from *squint* injected (red line) and Nodal inhibitor treated (blue line) embryos; peaks called by MACS program in *squint* injected sample; FoxH1 ChIP-Seq track from *squint* injected embryos; peaks called by MACS program in *squint* injected FoxH1 sample; genomic track.

ChIP analysis identified 1041 peaks Smad2 binding peaks in *squint* mRNA-injected embryos [Fig.3.1]. I identified 614 genes with TSS no further than 20kb away from a Smad2 peak. Gene Ontology analysis of these genes singled out these classes of genes as overrepresented: dorsal/ventral pattern formation; anterior/posterior axis specification; mesendoderm morphogenesis and cell migration; liver, digestive tract, pancreas, floor plate, and neural tube development; left/right asymmetry determination. Previously studied Nodal-regulated genes in vertebrates also regulate these classes of genes.

There was a strong overlap between Smad2 and FoxH1 binding sites [Fig.3.1], 58% of Smad2 peaks overlapped with a FoxH1 peak. I extracted genomic DNA sequence of Smad2 peaks and used XXMOTIF online tool to look for transcription factor binding elements (BEs). Consensus sequences for the following

transcription factors were found: FoxH1, Homeobox and Oct-family proteins, EomesA, and Smad2/3 [Fig.3.2]. FoxH1 transcription factor has long been known to be necessary for Nodal signaling (Pogoda et al., 2000; Weisberg et al., 1998) and so have homeobox factors Mixer and Mezzo (Kunwar et al., 2003; Poulain and Lepage, 2002). Oct-family proteins play an important role in the regulation of pluripotency in undifferentiated cells, and Oct1 transcription factor in particular had been shown to directly interact with Smad2 protein (Liu et al., 2011). EomesA is another transcription factor that has been implicated in this pathway in several model organisms, including zebrafish (Nelson et al., 2014). One study, using a dominant negative version of EomesA transcription factor in zebrafish, suggested that EomesA together with FoxH1 are the two main factors needed for the interpretation of Nodal signaling in embryonic cells (Slagle et al., 2011).

One of the particularly interesting motifs I identified is that of Osr1 transcription factor [Fig.3.2]. Odd skipped related 1 (Osr1) is a zinc finger transcription factor, homolog of the *Drosophila* pair-rule gene *odd*. Research in vertebrates showed that it is expressed in a distinctive way in the intermediate mesoderm, and plays an essential role in kidney formation in mice and zebrafish (Mudumana et al., 2008; Mugford et al., 2008). Loss of function studies in zebrafish revealed an additional role for *osr1* in endoderm patterning, as loss of *osr1* caused an increase in the number of cells expressing endoderm markers *sox17* and *foxa2* (Mudumana et al., 2008). Further studies showed that it is induced in mesendoderm by Nodal signaling, and that Osr1 activity is required to limit the number of *sox32* expressing endoderm cells (Terashima et al., 2014). Taken together, these results suggest that Osr1 is a negative regulator of endoderm-specific Nodal signaling. The mechanism of its inhibitory effect has not been studied, and our finding that Osr1

binding element is found in almost a quarter of Smad2 binding sites suggests that the inhibition might be achieved by competition for the binding sites. I looked for the genes that Smad2 peaks with an Osr1 binding element were close to, and identified 61 of those, one of which is *foxa3*, a definitive endoderm marker (Grapin-Botton, 2008). Gene Ontology analysis of these 61 genes showed that there was an enrichment of endoderm specification genes (p-value of 1.60E-02), but not mesodermal ones (Mi et al., 2013).

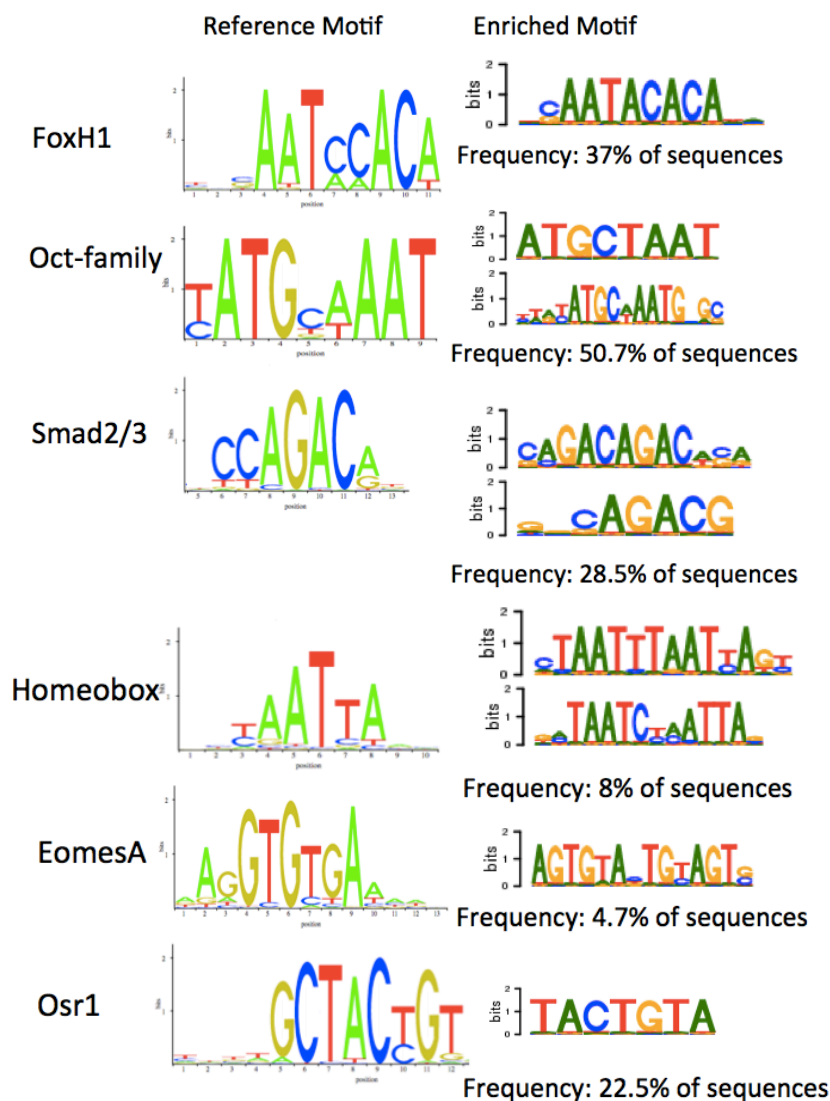


Figure 3.2

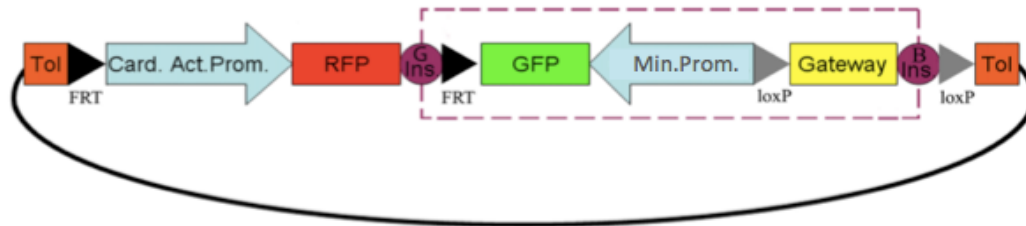
Binding motifs enriched in genomic sequences bound by Smad2 in Squint-injected embryos

FoxH1, EomesA, Smad2/3, Osr1, homeobox family and Oct-family binding sequences were enriched in the analyzed genomic sequences.

Identified regulatory elements can drive gene expression in a Nodal-responsive manner

To test whether the Smad2 binding loci are in fact functional enhancer elements, I created reporter constructs where these elements were driving the expression of *eGFP* and used them in transient transgenic fish experiments. I built reporter constructs, in which the tested DNA element was cloned in front of a minimal promoter followed by the *eGFP* transcript [Fig.3.3.A]. I used Zebrafish Enhancer Detector (ZED) vector that was designed specifically for testing regulatory elements in zebrafish embryos (Bessa et al., 2009). It contains a built-in minimal promoter for *eGFP*, which is unable to drive expression on its own, and a positive transgenic marker: RFP protein driven by cardiac actin promoter. This Tol2 transposon-based vector also contains insulator sequences to shield the minimal promoter from position effects (Bessa et al., 2009). I tested putative enhancers from 5 different known Nodal targets, in each case cloning an approximately 500bp DNA stretch centered on the Smad2 binding locus. The enhancers of *lhx1a*, *lft1*, *ntla*, *gsc* and *bhik* were tested in this manner. In all cases, this setup led to the expression of *eGFP* in the embryonic margin area at shield stage, which is where Nodal signaling is active in wild-type embryos [Fig.3.3.B-C].

A.

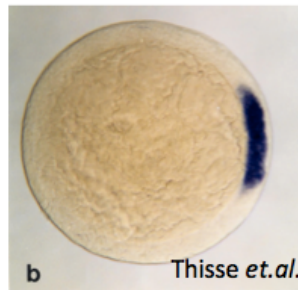
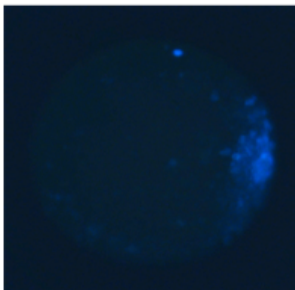


B. *gsc*

```

1  TGTTCGTCAGGTTTCATGAGTTGGATTGTGCACAATGCATTTTTTCTTCTGGTCAGTGATTGTTTATAATTATTAAGTGA
81  TATAAATTCATATAAATTTAAATCCGATTTTAAACAGATGATGAGTGCAGTGTGTAACACAAATATGACTGTTATTTATTC
161  ATTGTTTAGATATAATCAACCGTGCCATGCCACGGATATCCTCTGGTTTTCCATAAACAATTGAGATTACAAATGTGC
241  GATATTGGTGATAAATTGAATCGCCAATGTCACCGAGCAGCGTCTTCCTTCATGCAAATCCATTTGATCAAGCAATGTA
321  ATTGCAATAAGTATGTTTATAGAGCTGGAGAATAGGGCCGCTAATGGTCGTTATCTTATCACAGCGAGGGTCATATGTTTCG
401  TTGAGCGTGTGCTTATTGGAGGTGTTGTAAGTTAACACGCGCGAT

```



c. *lhx1a*

```

1  ATGGCTTTAAGGAAATCACACTTTGTTTACAACGAATGGCAGCCAGGGAAGCCGACGGGATCAATGCTTTGCCAGGGTAG
81  TTGAAAGTGTAACGTGGCCAGTGGCAGGCTGTCTGCACAACGCCGAGTGTGTTTGCAGTGGTTTGAGTTAAGATGTGAT
161  AAGTGGATTTCAATAATAAACGACTCTGTGATGGAGACCCTCAGACAAATGTAATCACAGACTCAGAAATCAACAATCTGCA
241  CAGAGAACCACAAATACACATTATTTGGTCATAATTAAGTGGTGAGAATTTACTGCAATGGTGTGTTGGATTGTAAGAATGAA
321  GATGCTACGGGAGATTATTAATTGTGCTTTAATGCACATTACTAACCCTGCTCAATAAACGTATATGTTTTTTAGGCCCTA
401  CAGTGGAAGGAAAGTACATGCTAATGCATATATAAATATTATTTGTTTGGCTATTCTAA

```

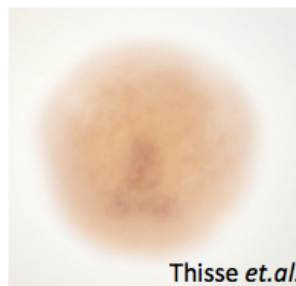
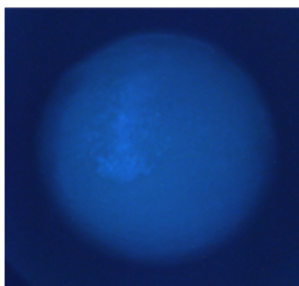
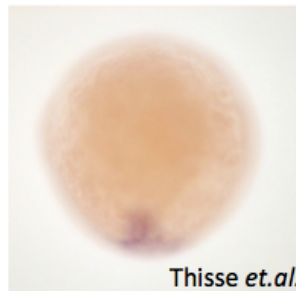
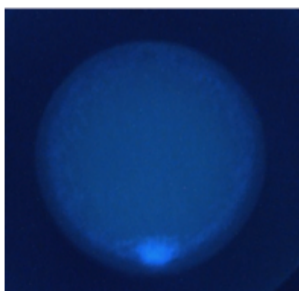


Figure 3.3 (continued on the next page)

Figure 3.3 (continued)

Smad2-bound putative enhancer elements drive gene expression in the embryo margin zone

A) Zebrafish Enhancer Detection (ZED) vector was used to test enhancer elements, as published in (Bessa et al., 2009). Gateway system was used to clone various putative enhancer elements into this vector.

B) Smad2-bound enhancer element 6kb upstream from the *gsc* gene TSS was used to drive *eGFP* expression in shield stage embryos. The sequence of the element shows consensus binding sequences of FoxH1 (cyan) and Smad2/3 (magenta). The picture on the left shows the observed *eGFP* expression pattern, while the picture on the right shows the expression pattern of the *gsc* gene visualized by *in situ* (from (Thisse and Thisse, 2008)).

C) The sequence of the enhancer element in the first intron of *lhx1a* gene, the observed *eGFP* expression driven by it and *in situ* pictures of *lhx1a* expression pattern (Thisse and Thisse, 2008) are shown.

To test whether the expression driven by these enhancers will be responsive to Nodal signaling levels, I treated the transfected embryos with Nodal-inhibiting drug, or injected them with 5pg of *squint* mRNA at one-cell stage. I tested enhancer elements from *ntla* and *lhx1a* genes, and this test showed that the information contained in that stretch of DNA is indeed sufficient to make gene expression Nodal-responsive [Fig.3.4].

A. *ntla*

```

1 | GTTCTGCAGGACGCTCATCAGGCTGGAGTCATGCTGGATTAAGGAAATTCACACAGATCGTCACCTCTGGCCATCCACAC
81 | TTCAGTCGCGTCAATGGTCGCATCCTAATGGCCGTCCTAGTAAACACGGGCTTCTCACATGTGTATCTGGGACAACAAA
161 | AGATTAGCATTATTCGCAATATACAACTGAGTTCAAAGGCAACGAGGCAGCAGACACCGTCTCTCTTCCTCTTTTGT
241 | TCAAAGACGAGCTCGCGAAGAGCAAAACATGTTTGGGTTTAGGCTTTCATTCCACAAATATCCAGTAAACAACCG

```

B.

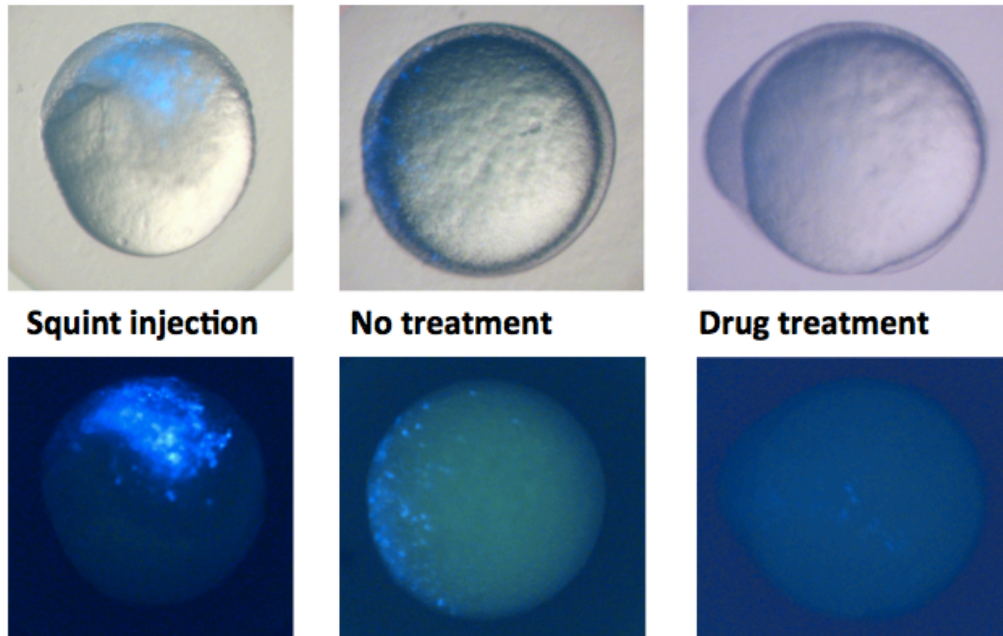


Figure 3.4

Expression driven by the regulatory element of ntl gene is responsive to Nodal signaling levels

A) Smad2-bound enhancer element 2kb upstream from the *ntl* gene TSS was used to drive *eGFP* expression in shield stage embryos. The sequence of the element shows consensus binding sequences of FoxH1 (cyan) and Smad2/3 (magenta).

B) From left to right, *eGFP* expression is shown in embryos injected with 5pg *squint* mRNA, untreated, or treated with Nodal drug inhibitor. Top row: bright-field view; bottom row: image taken by fluorescence microscope.

After exposure to low and high levels of Nodal signaling Smad2 binds near different groups of genes

Since the results of the first paper that published Smad2 ChIP in zebrafish showed very limited results (Liu et al., 2011), I directly compared my results to the paper published by the Wardle group in 2014 (Nelson et al., 2014). In this work they identified 897 Smad2 peaks and 561 genes within 20kb of these. My

ChIP-Seq analysis identified 1041 Smad2 peaks and 614 genes with 20kb. These two lists had 145 genes in common between them.

Transcription factor binding on its own is not sufficient to identify real targets of that factor, and gene expression data has to be used as additional evidence. To make sure that I have few false negatives, the list of potential Nodal targets I used was as comprehensive as possible. This list included 235 genes that were identified as Nodal-regulated in previously published literature (Bennett et al., 2007; Nelson et al., 2014), as well as 330 genes that were found to be Nodal-responsive using RNA-Seq data acquired in our lab [unpublished, Andrea Pauli]. Dr. Pauli collected gene expression data in dome stage zebrafish embryos that were unperturbed, injected with 5pg of *squint* mRNA, or were mutant for the *oep* gene (MZ*oep*). I chose the genes that, when compared to untreated embryos, were either upregulated by a factor of 1.5 in *squint* injected ones, or downregulated by a factor of 1.5 in MZ*oep* embryos. These two groups of Nodal targets genes had 83 targets in common, resulting in a list of 482 genes that change their expression in response to Nodal signaling, because of either direct or indirect effect on their transcription.

A gene was called a bona fide Nodal target if it both changed expression in response to a change in Nodal signaling levels, and had a Smad2 peak within 20kb of their TSS. There were 88 such genes based on the Wardle Smad2 dataset, and 111 based on mine, with 56 shared genes [Fig.3.5].

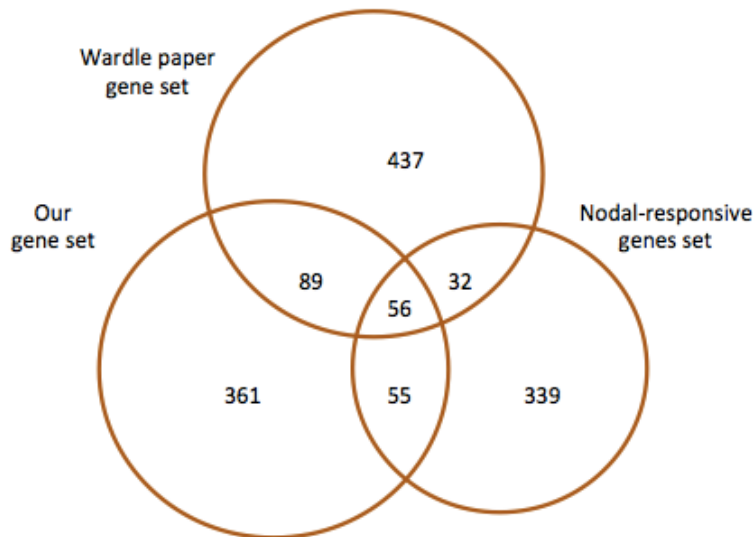


Figure 3.5

Overlap between Smad2-proximal genes identified based on the Wardle group data, my ChIP-Seq data, and Nodal-responsive genes

A large number of genes overlapped between these three groups. Nodal-responsive genes that do not have a Smad2 peaks in either of the ChIP-Seq datasets are probably not direct Nodal target genes. Of special interest are the two groups of genes that are Nodal responsive but have a proximal Smad2 peak only in one dataset or the other.

Table 3.1:

Previously unpublished Nodal target genes

acot11a	ptfla
cpn1	rasd1
dhrs3b	rfx2
dhrs9	rippy1
ednraa	s1pr5a
fam162a	sall3b
golim4b	samm50l
her6	six4b
her9	slc13a4
lmo1	slc25a22
mapk11	smtnl
mnx1	tcf3b
pcdh18a	uchl5

Our results therefore identified 55 direct Nodal target genes that were not picked up by the previously published research. Of those 55 genes, 29 were previously described as responsive to the levels of Nodal signaling (directly or indirectly), while 26 are previously unpublished novel target genes [Table 3.1]. Based on Gene Ontology analysis these genes were enriched for DNA-binding transcription factors, as well as genes involved in early developmental processes and anatomical structure development.

DISCUSSION

Smad2 is a factor that is indispensable to Nodal signaling transduction. Since its binding to DNA is considered to be the activating signal for target gene expression, we must understand its recruitment process to learn how cellular response to Nodal is regulated. This study gets us one step closer to this goal.

Understanding where and when Smad2 binds in zebrafish genome is of particular interest for developmental biology. It is more representative of actual embryonic development than studies done in cell culture, as we observe the regulation of signaling in its native context. Smad2 binding data can also be related to the *in vivo* Smad2 nuclear accumulation data to build better models of concentration-dependent target gene induction (Dubrulle et al., 2015).

Our study identified Smad2 binding peaks in embryos overexpressing Nodal signal Squint, and we found 55 Nodal-regulated genes not found in the previously published datasets (Liu et al., 2011; Nelson et al., 2014). To find out whether this difference is meaningful, one could compare the expression driven by the regulatory regions of some of these 55 genes and some of the genes that were only

identified in the Wardle group dataset. These regulatory regions could then also be tested for Nodal-responsiveness in zebrafish embryos. If the information in the sequence of these DNA elements is sufficient to dictate Smad2-binding conditions, we expect Smad2 to bind to the elements in the first group only when the cell is exposed to high Nodal levels. These enhancers should then drive expression only close to the margin of the embryo and in the shield region. Enhancers of genes that didn't have Smad2 binding near their TSS in our data, but had in the Wardle group data, should then drive expression further away from the margin.

Identifying the different factors that can bind to these elements could provide a lot of information on how gene expression threshold is determined. While it is possible that some of the threshold-determining information is contained in the promoter region rather than in the enhancer (Dubrulle et al., 2015), understanding Smad2 binding would undoubtedly provide us with some answers to threshold control. Another attractive approach to understating differential Smad2 binding to regulatory elements would be building synthetic enhancer elements. We could combine the consensus binding sequences of Smad2/3, FoxH1, EomesA, Oct1 and others in different numbers and positions to see how it affects expression.

We also identified 26 genes that were not previously shown to be Nodal-regulated. This group of genes showed enrichment for DNA-binding transcription factors and it brings us closer to building the gene regulatory network of vertebrate development.

Attempts have been made to find a small group (FoxH1 and Mixer; FoxH1 and eomesA) of transcription factors that are responsible for all Smad2-dependent Nodal signaling transduction in zebrafish, but a convincing result has yet to be reached (Kunwar et al., 2003; Slagle et al., 2011). It is likely that more than two

co-transcription factors have to be lost to faithfully mimic loss of Nodal signaling phenotype. A number of other factors, such as Mezzo, p53, Gata5 have also been shown to interact with Smad2 (Cordenonsi et al., 2003; Poulain and Lepage, 2002; Reiter et al., 2001). We also found the consensus binding sequence of Oct-family proteins in more than half of the Smad2 binding regions and we identified Osr1 as another potentially important regulator. Our study brings us one step closer to a comprehensive list of Smad2-interacting transcription factors.

I tested short DNA elements centered on a Smad2 binding peak for their ability to drive gene expression in a Nodal-dependent manner in transient transgenic fish. It would be very informative to do further experiments in stable transgenics, to find the precise region of their expression in wild-type embryos and embryos injected with various levels of *squint* mRNA. Another important experiment would be observing the behavior of these enhancers in FoxH1 mutant background. Any residual Nodal-dependent expression in such background should be the result of other Smad2 recruiting factors binding to it.

To sum up, the results I present in this chapter expand the repertoire of known Nodal target genes, identify an important group of Smad2-binding regulatory elements and test their responsiveness to Nodal signaling.

MATERIALS AND METHODS

Chromatin Immunoprecipitation (ChIP)

Zebrafish TLAB embryos were dechorionated with pronase at one-cell stage, and 300-500 embryos at dome stage were used for each experiment, as previously described (see Chapter II). Anti-Smad2 antibody (Invitrogen 51-1300) was used for the pull-down.

ChIP-qPCR

ChIP was done using shield stage embryos. The eluted DNA was analyzed using SYBR[®] Green I (BioRad) and DNA Engine Opticon System (MJ Research) with technical duplicates. ChIP enrichment was normalized to input, and shown as % input.

ChIP-Seq

ChIP with anti-Smad2 antibody was performed as described. Embryos were untreated, incubated at 1-cell stage with 10mM Nodal inhibitor drug SB-505124 (DaCosta Byfield et al., 2004) (S4696 SIGMA), or injected at 1-cell stage with 5pg of *squint* mRNA. Sample DNA was sequenced on Illumina Hi-Seq platform. ChIP-seq reads were mapped using Bowtie (v.0.12.9) (Langmead, 2010); peaks were called using MACS (v.2.0.10) (Zhang et al., 2008). De novo motif analyses used XXMOTIF (Hartmann et al., 2013). JASPAR public database was used to find candidate proteins. GO analyses were performed using PANTHER (protein annotation through evolutionary relationship) classification system (Mi et al., 2013).

RNA-Seq

Sample preparation was done as previously described (see Chapter II). Embryos were untreated, incubated at 1-cell stage with 10mM Nodal inhibitor drug SB-505124 (DaCosta Byfield et al., 2004) (S4696 SIGMA), or injected at 1-cell stage with 5pg of *squint* mRNA. Samples were sequenced on Illumina Hi-Seq platform, and aligned to *Danio rerio* genome Zv9 assembly. Differential gene expression analysis of RNA-Seq data was performed using TopHat (v. 1.3.3) (Trapnell et al., 2009) and Cuffdiff (v. 1.3.0) (Trapnell et al., 2012). GO analyses were performed using PANTHER (protein annotation through evolutionary relationship) classification system (Mi et al., 2013).

Transient transgenic zebrafish generation

Transposase (Kawakami et al., 2004) cDNA containing vector was linearized with NotI restriction enzyme and mRNA was transcribed using Sp6 RNA polymerase. The Tol2-based method of transgenesis (Kawakami et al., 2004) was used. At one-cell stage, each embryos was injected with a 1nl mixture containing 50 ng/ul of transposase mRNA and 40 ng/ul of purified vector DNA. Embryos were incubated until shield stage and then imaged.

Nodal induction dynamics (from (Dubrulle et al., 2015))

Wild-type embryos were first injected at the one-cell stage with 1 nl of *cyclops* and *squint* MOs mixture (at 0.2 mM and 4 µg/µl, respectively) to inhibit endogenous Nodal signals. Morphants were further injected with 0.5–1.5 nl of recombinant mouse Nodal protein (rmNodal, R&D Systems, Minneapolis, MN) at different concentrations in the extracellular space at 4hpf (sphere stage). For NanoString

analysis total RNA from 5 to 10 embryos for each data point was extracted using the RNeasy mini kit (Qiagen, The Netherlands) and 100ng of input RNA was processed through the nCounter assay using standard protocols (NanoString Technologies, Seattle, WA) (Kulkarni, 2011).

In situ hybridization

In situ hybridization on whole mount embryos was carried out using standard methods (Thisse and Thisse, 2008). Embryos were fixed in 4% formaldehyde overnight, and then dehydrated, rehydrated, and stained. Following staining, embryos were cleared in benzyl benzoate/benzyl alcohol (2:1 vol/vol) and imaged.

REFERENCES

- Bennett, J.T., Joubin, K., Cheng, S., Aanstad, P., Herwig, R., Clark, M., Lehrach, H., Schier, A.F., 2007. Nodal signaling activates differentiation genes during zebrafish gastrulation. *Dev. Biol.* 304, 525–540. doi:10.1016/j.ydbio.2007.01.012
- Bessa, J., Tena, J.J., De La Calle-Mustienes, E., Fernandez-Minan, A., Naranjo, S., Fernandez, A., Montoliu, L., Akalin, A., Lenhard, B., Casares, F., Gomez-Skarmeta, J.L., 2009. Zebrafish Enhancer Detection (ZED) vector: A new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. *Dev. Dyn.* 238, 2409–2417. doi:10.1002/dvdy.22051
- Brown, S., Teo, A., Pauklin, S., Hannan, N., Cho, C.H.-H., Lim, B., Vardy, L., Dunn, N.R., Trotter, M., Pedersen, R., Vallier, L., 2011. Activin/Nodal signaling controls divergent transcriptional networks in human embryonic stem cells and in endoderm progenitors. *Stem Cells* 29, 1176–1185. doi:10.1002/stem.666
- Chiu, W.T., Charney Le, R., Blitz, I.L., Fish, M.B., Li, Y., Biesinger, J., Xie, X., Cho, K.W.Y., 2014. Genome-wide view of TGF β /Foxh1 regulation of the early mesendoderm program. *Development* 141, 4537–47. doi:10.1242/dev.107227
- Cordenonsi, M., Dupont, S., Maretto, S., Insinga, A., Imbriano, C., Piccolo, S., 2003. Links between tumor suppressors: p53 is required for TGF- β gene responses by cooperating with Smads. *Cell* 113, 301–314. doi:10.1016/S0092-8674(03)00308-8
- DaCosta Byfield, S., Major, C., Laping, N.J., Roberts, A.B., 2004. SB-505124 is a selective inhibitor of transforming growth factor-beta type I receptors ALK4, ALK5, and ALK7. *Mol. Pharmacol.* 65, 744–752. doi:10.1124/mol.65.3.744
- Dubrulle, J., Jordan, B.M., Akhmetova, L., Farrell, J.A., Kim, S.-H., Solnica-Krezel, L., Schier, A.F., 2015. Response to Nodal morphogen gradient is determined by the kinetics of target gene induction. *Elife* 4. doi:10.7554/eLife.05042
- Fei, T., Zhu, S., Xia, K., Zhang, J., Li, Z., Han, J.-D.J., Chen, Y.-G., 2010. Smad2 mediates Activin/Nodal signaling in mesendoderm differentiation of mouse embryonic stem cells. *Cell Res.* 20, 1306–1318. doi:10.1038/cr.2010.158
- Grapin-Botton, A., 2008. Endoderm specification. *StemBook* 3–4. doi:10.3824/stembook.1.30.1
- Hartmann, H., Guthohrlein, E.W., Siebert, M., Luehr, S., Soding, J., 2013. P-value-based regulatory motif discovery using positional weight matrices. *Genome Res.* 23, 181–194. doi:10.1101/gr.139881.112
- Kawakami, K., Takeda, H., Kawakami, N., Kobayashi, M., Matsuda, N., Mishina, M., 2004. A transposon-mediated gene trap approach identifies developmentally

- regulated genes in zebrafish. *Dev. Cell* 7, 133–144.
doi:10.1016/j.devcel.2004.06.005
- Kim, S.W., Yoon, S.J., Chuong, E., Oyolu, C., Wills, A.E., Gupta, R., Baker, J., 2011. Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. *Dev. Biol.* 357, 492–504.
- Kulkarni, M.M., 2011. Digital multiplexed gene expression analysis using the nanostring ncounter system. *Curr. Protoc. Mol. Biol.*
doi:10.1002/0471142727.mb25b10s94
- Kunwar, P.S., Zimmerman, S., Bennett, J.T., Chen, Y., Whitman, M., Schier, A.F., 2003. Mixer/Bon and FoxH1/Sur have overlapping and divergent roles in Nodal signaling and mesendoderm induction. *Development* 130, 5589–5599.
doi:10.1242/dev.00803
- Langmead, B., 2010. Aligning short sequencing reads with Bowtie. *Curr. Protoc. Bioinforma.* doi:10.1002/0471250953.bi1107s32
- Liu, Z., Lin, X., Cai, Z., Zhang, Z., Han, C., Jia, S., Meng, A., Wang, Q., 2011. Global identification of SMAD2 target genes reveals a role for multiple co-regulatory factors in zebrafish early gastrulas. *J. Biol. Chem.* 286, 28520–28532.
doi:10.1074/jbc.M111.236307
- Mi, H., Muruganujan, A., Casagrande, J.T., Thomas, P.D., 2013. Large-scale gene function analysis with the PANTHER classification system. *Nat. Protoc.* 8, 1551–66. doi:10.1038/nprot.2013.092
- Mudumana, S.P., Hentschel, D., Liu, Y., Vasilyev, A., Drummond, I. a, 2008. Odd Skipped Related1 Reveals a Novel Role for Endoderm in Regulating Kidney Versus Vascular Cell Fate. *Development* 135, 3355–67. doi:10.1242/dev.022830
- Mugford, J.W., Sipilä, P., McMahon, J.A., McMahon, A.P., 2008. Osr1 expression demarcates a multi-potent population of intermediate mesoderm that undergoes progressive restriction to an Osr1-dependent nephron progenitor compartment within the mammalian kidney. *Dev. Biol.* 324, 88–98.
doi:10.1016/j.ydbio.2008.09.010
- Nelson, A.C., Cutty, S.J., Niini, M., Stemple, D.L., Flicek, P., Houart, C., Bruce, A.E.E., Wardle, F.C., 2014. Global identification of Smad2 and Eomesodermin targets in zebrafish identifies a conserved transcriptional network in mesendoderm and a novel role for Eomesodermin in repression of ectodermal gene expression. *BMC Biol.* 12, 81. doi:10.1186/s12915-014-0081-5
- Pogoda, H.M., Solnica-Krezel, L., Driever, W., Meyer, D., 2000. The zebrafish forkhead transcription factor FoxH1/Fast1 is a modulator of Nodal signaling required for organizer formation. *Curr. Biol.* 10, 1041–1049. doi:10.1016/S0960-9822(00)00669-2

- Poulain, M., Lepage, T., 2002. Mezzo, a paired-like homeobox protein is an immediate target of Nodal signalling and regulates endoderm specification in zebrafish. *Development* 129, 4901–4914.
- Reiter, J.F., Kikuchi, Y., Stainier, D., 2001. Multiple roles for Gata5 in zebrafish endoderm formation. *Development* 128, 125–35.
- Slagle, C.E., Aoki, T., Burdine, R.D., 2011. Nodal-dependent mesendoderm specification requires the combinatorial activities of FoxH1 and eomesodermin. *PLoS Genet.* 7. doi:10.1371/journal.pgen.1002072
- Terashima, A. V., Mudumana, S.P., Drummond, I.A., 2014. Odd skipped related 1 is a negative feedback regulator of nodal-induced endoderm development. *Dev. Dyn.* 243, 1571–1580. doi:10.1002/dvdy.24191
- Thisse, C., Thisse, B., 2008. High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat. Protoc.* 3, 59–69. doi:10.1038/nprot.2007.514
- Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111. doi:10.1093/bioinformatics/btp120
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., Pachter, L., 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–78. doi:10.1038/nprot.2012.016
- Weisberg, E., Winnier, G.E., Chen, X., Farnsworth, C.L., Hogan, B.L.H., Whitman, M., 1998. A mouse homologue of FAST-1 transduces TGF β superfamily signals and is expressed during early embryogenesis. *Mech. Dev.* 79, 17–27. doi:10.1016/S0925-4773(98)00160-9
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., Liu, X.S., 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137. doi:10.1186/gb-2008-9-9-r137

CHAPTER IV

Analysis of the relationship between Nodal target gene expression and histone modifications

ABSTRACT

While it has been demonstrated in many systems that chromatin marks are strongly associated with the transcriptional status of genes, many questions in this area still remain unanswered. In this chapter I describe the study of the interaction between Nodal signaling in early zebrafish development and chromatin marks. I found that exposure of embryonic cells to high levels of Nodal is associated with low levels of H3K27me3 and high levels of H3K4me3 marks on Nodal target genes, relative to those in unexposed cells. One of the most important questions in the field of epigenetics is the causal relationship between gene transcription and histone marks, as well as the chromatin-dependent mechanisms by which transcription factors modulate target gene expression. Here I describe a Cas9-based system that we used to change H3K27me3 levels in a targeted manner, and tested it in a developing embryo on a Nodal-responsive *fgf8a* gene. Our results suggest that reduction of H3K27me3 mark on its own is not sufficient to affect the expression of this gene, and additional mechanisms are involved in target gene activation by Smad2.

PREFACE

The experiments were designed, performed, and interpreted by Summer Thyme and me, with equal contribution. I analyzed the H3K4me3 and H3K27me3 ChIP-Seq data generated by Nadine Vastenhouw. Andrea Pauli generated RNA-Seq data in *MZoep* and *squint* mRNA-injected zebrafish. Alexander F. Schier helped conceive, design, and support the project.

INTRODUCTION

Nodal signaling pathway drives the differentiation of pluripotent cells toward mesendodermal lineage. It has not, however, been extensively investigated how this signaling network cooperates with chromatin state during embryonic development. Zebrafish embryonic genome is inactive until maternal-zygotic transition, when transcription is initiated (Schier, 2007). After genome activation many genes acquire either an H3K4me3 or H3K27me3 chromatin mark at their promoters, which are associated with active and silenced genes, respectively (Vastenhouw et al., 2010). Some developmentally important genes, including a number of Nodal targets, are marked by bivalent chromatin domains, with both H3K4me3 and H3K27me3 marks present at their promoter. In embryonic stem cells these bivalent domains have been hypothesized to ‘poise’ developmental genes for rapid activation (Bernstein et al., 2006) and then resolve into either repressive (H3K27me3) or active (H3K4me3) state upon differentiation (Mikkelsen et al., 2007; Pan et al., 2007).

A study in human embryonic stem cells (hESCs) examined how Nodal signaling interacts with and affects particular chromatin states during endoderm commitment (Kim et al., 2011). This study showed that exposure of hESCs

to Nodal signaling leads to the activation of Nodal target genes and the concomitant reduction of H3K27me3 and increase of H3K4me3 marks at their promoters, in a manner highly correlated with SMAD2/3 binding. A mechanism for this process has been proposed: it has been demonstrated that SMAD2/3 was capable of recruiting the histone demethylase, JMJD3, to the NODAL promoter in mouse ESCs, causing the loss of the repressive mark H3K27me3 (Dahle et al., 2010). This study suggested that Nodal signaling induces target gene expression primarily through de-repression caused by the removal of H3K27me3 mark.

It is not known how H3K4me3 and H3K27me3 marks on Nodal target genes behave in vivo when undifferentiated cells are exposed to Nodal signaling. To address this question, we combined whole-genome H3K27me3 and H3K4me3 ChIP-Seq and targeted chromatin modification at the Nodal-responsive gene *fgf8a* in zebrafish embryos. We found that exposure to high levels of Nodal is associated with low levels of H3K27me3 and high levels of H3K4me3 marks on Nodal target genes. Our results also suggest that reduced H3K27me3 is associated with but not sufficient for Nodal target gene activation.

RESULTS

Nodal signaling increases H3K4me3 and reduces H3K27me3 marks at target genes

To determine the H3K4me3 and H3K27me3 states of Nodal-regulated genes, we employed genome-wide ChIP-Seq in zebrafish embryos at shield stage (6 hpf). In wild-type embryos Nodal signaling at this stage is fully active and gastrulation has begun. Under normal conditions only a subset of cells close to the

margin of the embryo is exposed to Nodal signaling, creating a heterogeneous cell population. To compare homogeneous cell populations, we used embryos in which Nodal signals were either expressed ubiquitously or globally repressed. Ubiquitous expression was achieved by injecting 5 pg of *squint* mRNA at one-cell stage (Feldman et al., 1998). To achieve global inhibition of the signaling, a drug that specifically blocks Nodal receptors, SB-505124 (DaCosta Byfield et al., 2004), was added to the medium in which embryos were incubated. In order to identify regions that exhibit a change in H3K27me3 or H3K4me3 levels in response to Nodal signaling, we used MACS peak-calling program on the acquired ChIP-Seq data (Zhang et al., 2008b).

H3K4me3 and H3K27me3 marks on important developmental genes are subject to extensive changes during early development. As we wanted to understand the dynamics of these marks on Nodal-regulated genes in particular, we focused on a short list of genes that were defined as direct targets using stringent criteria. A gene had to exhibit increased Nodal-induced expression visible by *in situ* hybridization, could be induced by Nodal in the presence of translation-blocking cycloheximide, and be bound by Smad2 and FoxH1 transcription factors (Dubrulle et al., 2015). Not all genes from this list are expressed at shield stage, so we isolated and sequenced mRNA from shield stage embryos injected with 5 pg of *squint* mRNA and MZ*oep* mutants. Based on the collected RNA-Seq data, we focused on 24 genes that exhibited more than a threefold change in expression between these two conditions at shield stage.

We then looked at the levels of H3K4me3 and H3K27me3 mark at the TSS of these genes at shield stage. Of these 24 genes, 18 exhibited higher levels of H3K4me3 mark and 12 exhibited lower levels of H3K27me3 mark in response to

Nodal signaling, with 10 genes displaying a change in both marks [Fig.4.1]. In total among 16130 RefSeq genes we looked at, 344 showed a change in H3K4me3 levels, and 45 in H3K27me3 levels. In some cases, Nodal-regulated genes completely lacked H3K27me3 marks in their promoter regions in the *squint*-injected sample. For example, the Nodal target gene *fgf8a* was associated with very high H3K4me3 and low H3K27me3 levels in embryos injected with *squint* mRNA [Fig.4.1]. These results indicate that Nodal signaling increases H3K4me3 and reduces H3K27me3 mark at the majority of its direct targets.

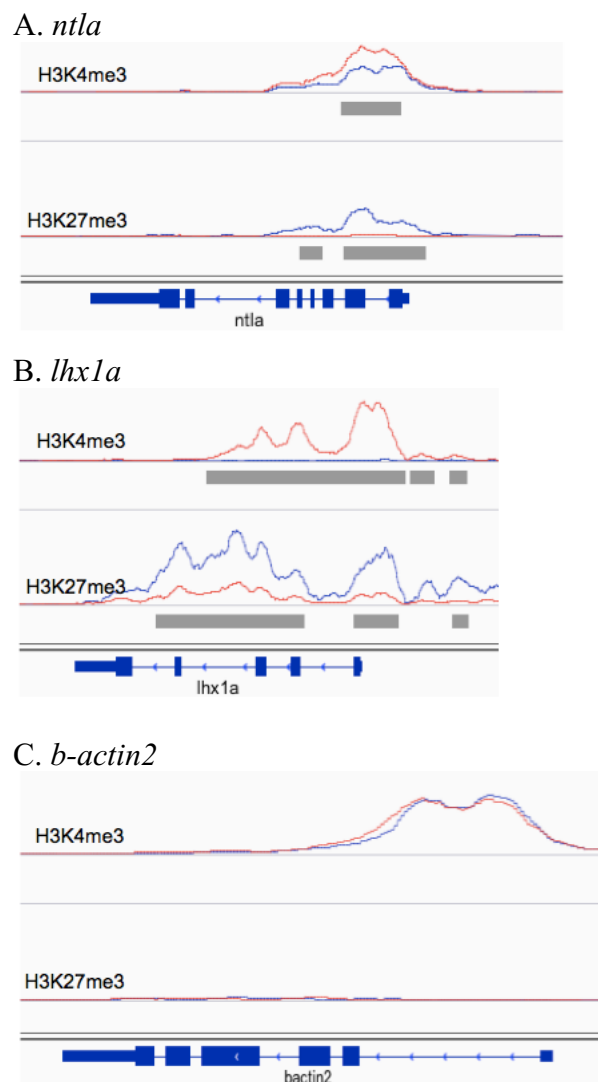
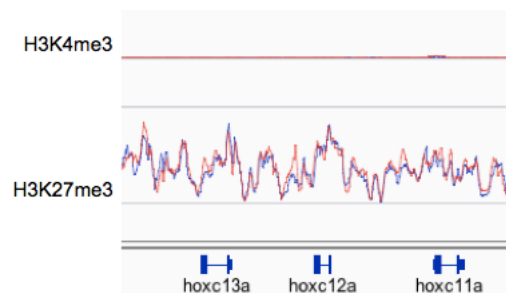


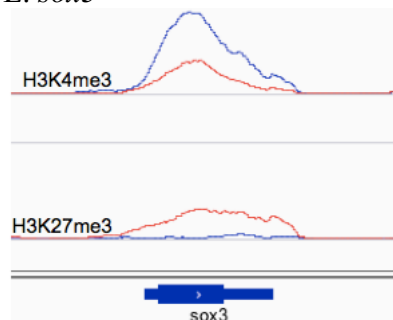
Figure 4.1 (continued on the next page)

Figure 4.1 (continued)

D. *hox* gene cluster



E. *sox3*



Legend

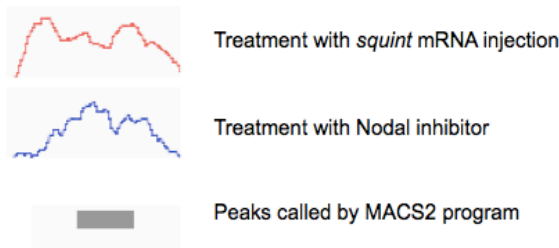


Figure 4.1

Genome-wide chromatin modification changes in zebrafish embryo in response to Nodal signaling

Examples of chromatin mark change on different genes in response to Nodal signaling as demonstrated by ChIP-Seq at shield stage. Peak tracks show the peaks called by MACS2 using one treatment as the tag track, and the opposite treatment as the control track.

A - B) *ntla* and *lhx1a* are well-studied Nodal signaling targets. In response to Nodal signaling (red line) both exhibit a higher level of H3K4me3 and lower level of H3K27me3.

C) *b-actin2* is a housekeeping gene, which is not affected by Nodal levels. Under each of the two conditions it has high H3K4me3 levels and no H3K27me3 mark.

D) *hox11a*, *hox12a* and *hox13a* genes are silenced at shield stage and have only H3K27me3 mark.

E) *sox3* is a gene that is inhibited by Nodal signaling. The levels of both histone marks behave in manner opposite to the positively regulated Nodal target genes.

Synthetic activation of a Nodal target gene

In order to study the interplay between chromatin modifications and gene expression in more detail, we focused on the Nodal-regulated gene *fgf8a*, as this gene was a particularly good candidate for our study. *fgf8a* has a very strong H3K4me3 peak and almost no H3K27me3 in *squint* injected embryos, and the reverse of it in embryos treated with a Nodal signaling inhibitor [Fig.4.2.A].

We first asked whether the changes in H3K4me3 and H3K27me3 marks are specific to Nodal-induced overactivation of the gene or whether they are a property of *fgf8a* activation in general. To distinguish between these possibilities, we targeted a fusion construct of catalytically dead Cas9 (dCas9) with the VP64 activation domain to the promoter region of *fgf8a* [Fig.4.2.A and B]. Targeting dCas9-VP64 to two gRNA binding sites upstream of the transcription start site increased *fgf8a* expression [Fig4.2.C].

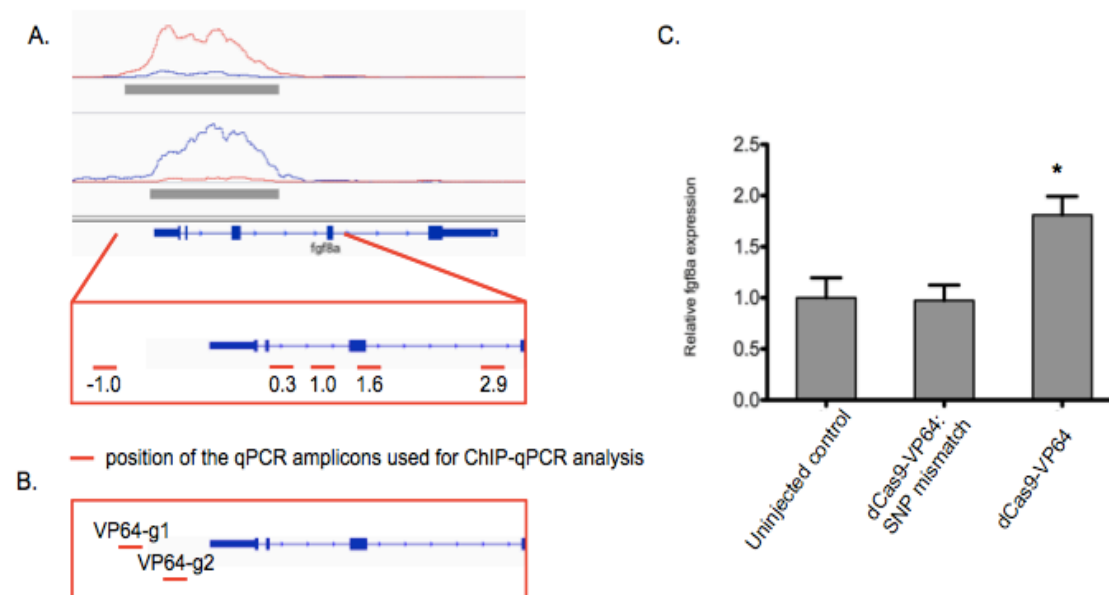


Figure 4.2 (continued on the next page)

Figure 4.2 (continued)

Induced gene expression using targeted dCas9-VP64 localization

(A) ChIP-Seq tracks of H3K27me3 and H3K4me3 histone marks on *fgf8a* promoter in wild type shield stage zebrafish embryos. Insert: qPCR amplicons used to analyze H3K27me3 and H3K4me3 marks by ChIP-qPCR, tagged by the distance from *fgf8a* transcription start site (TSS) in kbs.

(B) Positions of the two gRNAs used for dCas9-VP64 targeting to the *fgf8a* promoter.

(C) dCas9-VP64 increases *fgf8a* expression compared to uninjected control. In fish strain with a SNP mismatch (dCas9-VP64 SNP mismatch) in its genome, no increase is observed.

Targeting dCas9-VP64 to the *fgf8a* transcription start site also reduced H3K27me3 and increased H3K4me3 marks [Fig.4.3]. The effects on *fgf8a* expression and H3K4 and H3K27 methylation were completely abrogated [Fig.4.2.C] when dCas9-VP64 was targeted to an *fgf8a* allele that had a single mismatch within the target site of one of the gRNAs. These results showcase the use of naturally occurring SNPs as experimental controls for Cas9-based tools and reveal that a synthetic activator can induce transcription and changes in H3K4 and H3K27 methylation at a Nodal target gene.

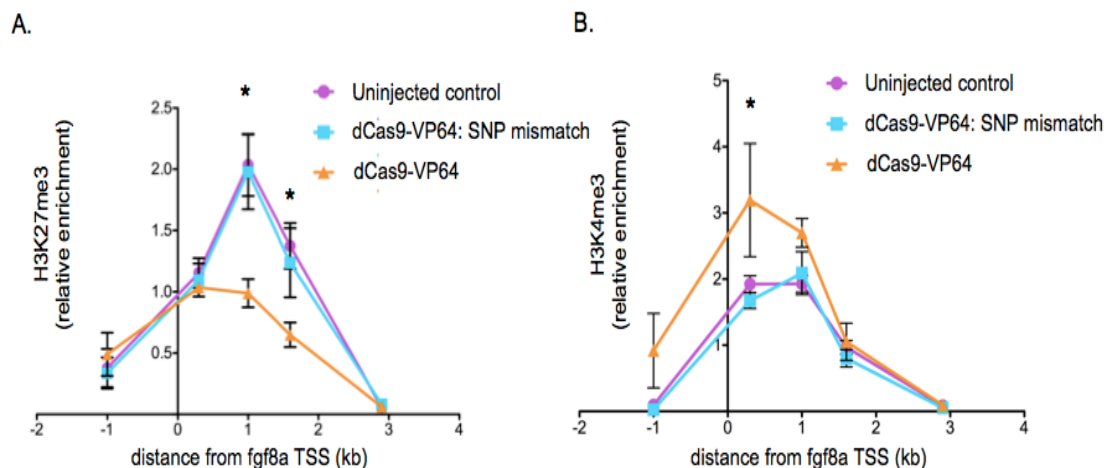


Figure 4.3

Changes in chromatin marks induced by targeted dCas9-VP64 localization

ChIP-qPCR of H3K27me3 (A) and H3K4me3 (B) mark on *fgf8a* gene shows a decrease in the levels of the mark of the former and increase in the levels of the latter following dCas9-VP64 injection. No change in the levels of either mark is observed in the fish strain with a SNP mismatch.

Targeted reduction of H3K27me3 levels

Cell culture studies of a Nodal target gene suggested that reduction of H3K27me3 marks might be sufficient for gene activation (Dahle et al., 2010). To test if this hypothesis holds true in vivo, we wished to specifically reduce H3K27me3 at the *fgf8a* locus. Standard approaches, such as histone modifier mutants, affect chromatin and gene expression globally, making it impossible to conclusively test the role of a specific chromatin mark at individual genes. For a locus-specific effect, sequence-specific DNA-binding domains or Cas9-based fusions can instead be used to locally change DNA methylation or histone marks (De Groote et al., 2012; Hilton et al., 2015; Kearns et al., 2015; Keung et al., 2014; Mendenhall et al., 2013; Rush et al., 2009). We therefore decided to recruit the H3K27me3 demethylase domain of UTX fused to dCas9 (dCas9-UTX) to the *fgf8a* locus [Fig.4.4.A]. To control for changes in H3K27me3 not due to UTX catalytic activity, we also generated a catalytically inactive version of the protein (dCas9-UTX-H1146A). Six gRNAs were designed to target the H3K27me3-containing region of *fgf8a* [Fig.4.4.A]. dCas9-UTX, but not dCas9-UTX-H1146A, reduced H3K27me3 across the *fgf8a* gene [Fig.4.4.B]. No change was detected for H3K4me3 levels [Fig.4.4.C] and nucleosome occupancy at the *fgf8a* locus, and H3K4me3 and H3K27me3 marks at other Nodal target genes were not altered. These results demonstrate that dCas9-UTX can reduce H3K27me3 levels in vivo at a locus of choice.

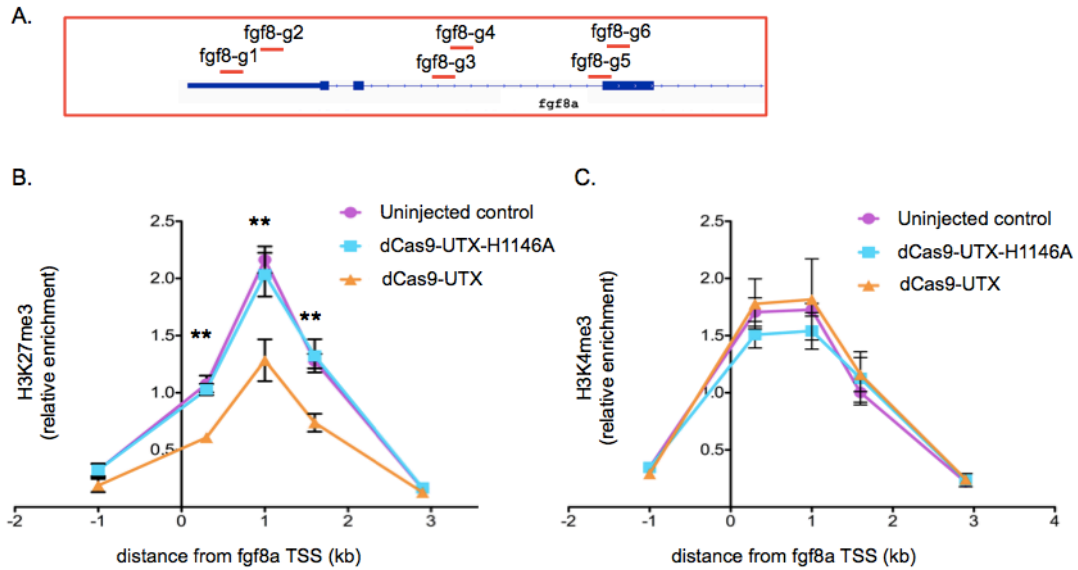


Figure 4.4

H3K7me3 mark reduction by dCas9-UTX

(A) Positions of the six gRNAs used for dCas9-UTX targeting to *fgf8a*.

ChIP-qPCR of H3K27me3 (B) and H3K4me3 (C) mark on *fgf8a* gene shows a decrease in levels of H3K27me3, but no change of H3K4me3. No change in the levels of either mark is observed with the catalytically inactive version of UTX: dCas9-UTX-H1146A.

To determine whether the six gRNAs contributed differentially to H3K27me3 reduction, we analyzed subsets of the guide RNAs. gRNAs fgf8-g3 and fgf8-g4 were tested individually, and three different pairs of gRNAs were analyzed (fgf8-g1/g2, fgf8-g5/g6, fgf8-g2/g5) [Fig.4.5.A]. None of these subsets reduced the H3K27me3 mark, while a combination of 4 gRNAs (fgf8-g1/g2/g5/g6) reduced H3K27me3 to similar levels as all six guide RNAs [Fig.4.5.B]. The changes in H3K27me3 levels achieved by dCas9-UTX with either 4 or 6 gRNAs were similar to those achieved by dCas9-VP64. These results show that combinations of 4 gRNAs are sufficient to target dCas9-UTX to a specific locus and induce the local reduction of H3K27me3.

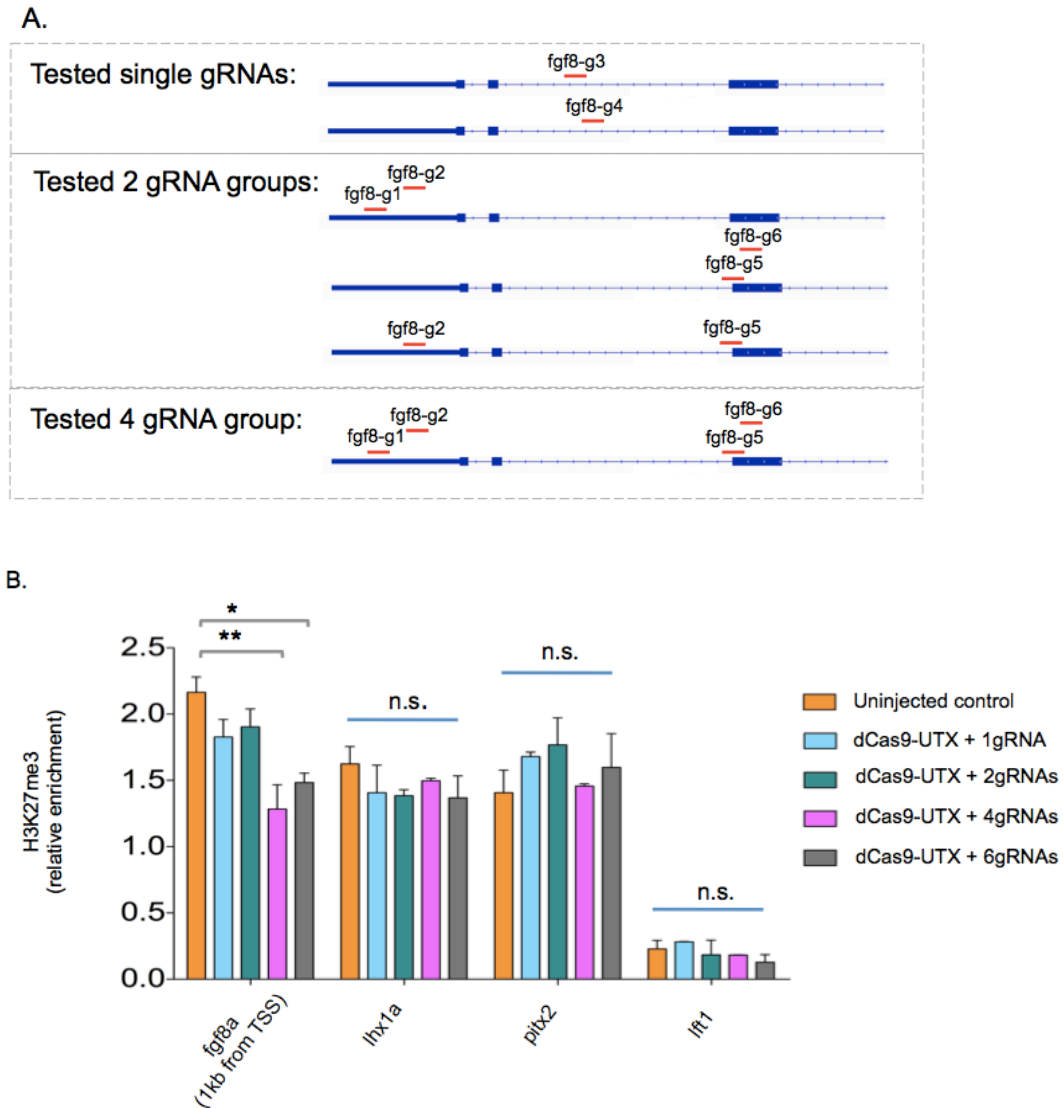


Figure 4.5

Combinations of 4 gRNAs are sufficient for H3K7me3 mark reduction

(A) Different gRNA combinations were coinjected with dCas9-UTX: 1 gRNA only (either *fgf8*-g3 or *fgf8*-g4); 2 gRNAs together ((*fgf8*-g1/*g2*) or (*fgf8*-g5/*g6*) or (*fgf8*-g2/*g5*)); 4 gRNAs together (*fgf8*-g1/*g2*/*g5*/*g6*).

(B) ChIP-qPCR measurement of H3K27me3 mark on *fgf8a* gene and 3 negative control regions (*lhx1a*, *pitx2*, *lft1*) upon the injection of different combinations of gRNAs. Combination of 4 gRNAs was equally as effective as 6 gRNAs, while no other combination changed H3K27me3 levels.

Reduction of H3K27me3 is not sufficient to increase target gene expression

We analyzed *fgf8a* mRNA levels in embryos injected with 4 gRNAs and the dCas9-UTX fusion to determine whether reduction of H3K27me3 levels at the *fgf8a* locus affects gene expression. As a control, we injected a dCas9-sfGFP fusion that does not affect H3K27me3 levels. No difference in expression was detected either by RT-qPCR or *in situ* hybridization [Fig.4.6.A and B], suggesting that reduction in H3K27me3 levels alone is not sufficient to cause an increase in the expression of this Nodal target gene.

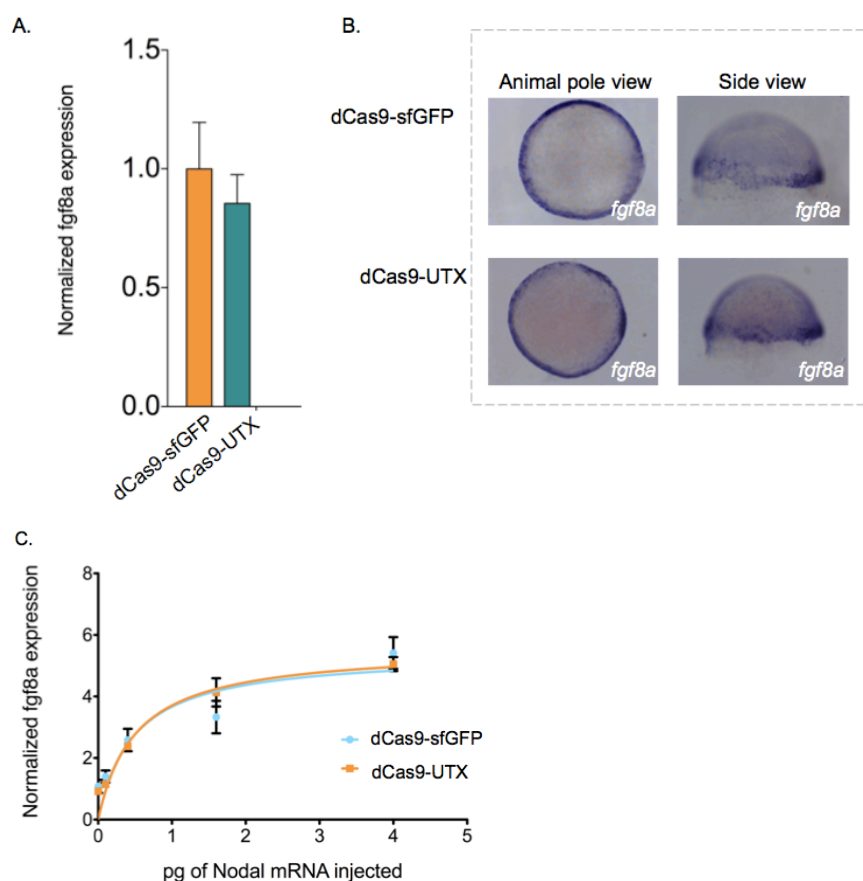


Figure 4.6

fgf8a gene expression levels following the injection of dCas9-sfGFP or dCas9-UTX construct with 4 gRNAs

(A) RT-qPCR measurements of *fgf8a* expression with the injection of dCas9-UTX or dCas9-sfGFP showed no significant difference.

(B) Whole mount *in situ* hybridization of the injected embryos at shield stage showed no increased or ectopic expression of *fgf8a*.

(C) Response of *fgf8a* to Nodal overexpression, as achieved by *squint* mRNA injection, was similar.

To test whether the responsiveness of *fgf8a* to Nodal signaling might be affected by reduction of H3K27me3 levels, Nodal signaling was increased by injection of different amounts of *squint* mRNA or decreased by addition of the Nodal receptor inhibitor SB-505124 in the presence of dCas9-UTX or dCas9-sfGFP fusions. The levels of expression, as measured by RT-qPCR, were in all cases similar between the embryos injected with dCas9-UTX and dCas9-sfGFP constructs [Fig.4.6.C]. These results suggest that H3K27me3 reduction is not sufficient to render a target gene more responsive to Nodal signaling.

DISCUSSION

Our study of histone marks at Nodal target genes generates tools and datasets for the analyses of chromatin modifications during early embryogenesis and indicates that the reduction of H3K27me3 levels is associated with but not sufficient for gene activation. ChIP-seq analyses reveal that Nodal signaling affects the chromatin landscape of its target genes: the repressive H3K27me3 mark decreases on target gene promoters, while the mark associated with active genes, H3K4me3, increases. While we find that such changes accompany both Nodal-induced and dCas9-VP64-induced gene expression, simply reducing H3K27me3 levels by dCas9-UTX is not sufficient for gene activation.

Both short and long-term effects of Nodal signaling on chromatin have been previously studied in human ES cells (hESCs). A recent study showed that a short (2-8 hours) treatment of hESCs with Nodal signaling inhibitor led to a rapid decrease in H3K4me3 levels, but did not affect H3K27me3 levels on any of the target genes (Bertero et al., 2015). This result suggests a model where Nodal signaling is

responsible for the deposition/maintenance of the activating mark, but does not directly affect the levels of H3K27me3. In contrast, the results of the ChIP-Seq experiments we performed show that in a developing embryo the levels of both of these marks are affected by Nodal signaling. Another study in hESCs showed that endodermal cells derived after a 5-day treatment with Activin exhibited a marked change in the levels of both H3K4me3 and H3K27me3 histone marks (Kim et al., 2011). Similar results were observed in mouse ES cells, and suggested that the key role of Nodal signaling was to recruit H3K27me3 demethylases to target genes to de-repress gene expression (Dahle et al., 2010). In this scenario, the default state of Nodal target gene is to be active unless repressed by H3K27me3 marks. Our results, however, show that local reduction of H3K27me3 levels by dCas9-UTX is not sufficient to upregulate *fgf8a* expression in vivo or to increase its responsiveness to activating signals. This finding suggests that Nodal signaling activates its target genes by regulating more than the deposition and removal of H3K27me3 marks. Although it is conceivable that other Nodal target genes respond differently or that removal of additional repressive marks would activate *fgf8a*, we favor a model in which Nodal signaling not only de-represses target genes but also actively switches them on. In this model, Nodal-induced phosphorylation of Smad2 (Dubrulle et al., 2015), recruits transcription factors, activating chromatin modifiers, and RNA polymerase to cis-regulatory elements to promote transcription. Notably, the dCas9-UTX fusion protein reduced H3K27me3 marks only when targeted to several sites within the target locus, suggesting that this fusion protein acts locally or that synergistic interactions are needed to demethylate histone marks in vivo.

Our study provides the first application of targeted histone modifying technologies to a vertebrate embryo. dCas9-effector fusions have been previously

used in tissue culture to study the effect of H3K27ac and H3K4me2 levels on enhancer activity (Hilton et al., 2015; Kearns et al., 2015). Using targeted recruitment of additional histone modifiers, the role of chromatin marks can now be interrogated directly, rather than through correlations in mutants. Such tools will be indispensable in dissecting the complex relationship between chromatin marks and gene regulation during vertebrate development.

MATERIALS AND METHODS

Chromatin Immunoprecipitation (ChIP)

Zebrafish TLAB embryos were dechorionated with pronase at one-cell stage, and 300-500 embryos at shield stage were used for each treatment, as previously described (see Chapter II). Antibodies used were histone H3 (Abcam ab1791), H3K4me3 (Millipore 07-473), H3K27me3 (Millipore 07-449).

ChIP-qPCR

ChIP was done using shield stage embryos. The eluted DNA was analyzed using SYBR[®] Green I (BioRad) and DNA Engine Opticon System (MJ Research) with technical duplicates. ChIP enrichment was normalized to input, and shown as enrichment relative to an intergenic control region upstream of gene *lft2*. Statistical analysis was performed with the Student's t test on paired data sets to calculate two-tailed P values.

ChIP-Seq

ChIP with H3K4me3 and H3K27me3 antibody was performed as described. Embryos were untreated, incubated at 1-cell stage with 10mM Nodal inhibitor drug SB-505124 (DaCosta Byfield et al., 2004)(S4696 SIGMA), or injected at 1-cell stage with 5pg of *squint* mRNA. Sample DNA was sequenced on Illumina Hi-Seq platform, and aligned to *Danio rerio* genome Zv9 assembly. For the analysis of Nodal responsive chromatin marks, peaks were called using MACS2 program (Zhang et al., 2008a) (default settings, except q-value = 0.005), comparing the *squint* injected and drug treated samples. To identify regions where H3K4me3 levels increased in response to Nodal

signaling, *Squint*-injected samples were used as tag tracks, and the drug-treated samples as the control tracks. The reverse was done to find regions with increased levels of H3K27me3 upon Nodal depletion: drug-treated sample was used as the tag track, *squint*-injected sample as the control track.

RNA-Seq

Sample preparation was done as previously described (see Chapter II). Embryos were untreated, incubated at 1-cell stage with 10mM Nodal inhibitor drug SB-505124 (DaCosta Byfield et al., 2004)(S4696 SIGMA), or injected at 1-cell stage with 5pg of *squint* mRNA. Samples were sequenced on Illumina Hi-Seq platform, and aligned to *Danio rerio* genome Zv9 assembly.

Plasmid constructs

All dCas9 fusion constructs – dCas9-VP64, dCas9-sfGFP (super-folder GFP (Pédelacq et al., 2006)), dCas9-UTX, and dCas9-UTX-H1146A – were cloned into the pCS2+ vector using isothermal assembly (Gibson et al., 2009). The sequence of UTX was derived from the zebrafish *kdm6al* gene (previously name *utx1*) by amplifying from shield stage cDNA. The UTX protein was truncated to a catalytic jumonji domain that has been crystallized for the human coding sequence (Hong et al., 2007), which has 78% identity to the same domain in zebrafish *kdm6al*. A 20 amino acid linker was used to connect dCas9 to UTX and sfGFP, while a linker of five amino acids was used in dCas9-VP64. The catalytic activity of UTX was eliminated by mutating a conserved histidine residue, number 1146 in the human UTX protein to alanine in the dCas9-UTX-H1146A construct. All constructs were linearized with NotI, and mRNA was transcribed using the mMachine SP6 kit

(Ambion), DNase treated, and precipitated with lithium chloride. RNA concentration was quantified using Nanodrop spectrophotometer.

Production and testing of gRNAs

Templates for gRNA transcription were generated by annealing oligonucleotides containing the T7 promoter and a 20 base-pair gene-specific sequence to an oligonucleotide encoding the reverse complement of the constant region of the gRNA, as previously described (Gagnon et al., 2014). The annealed product was converted to a double-stranded template by filling in the ssDNA overhangs with T4 DNA polymerase (NEB) and purified with the E.Z.N.A. Cycle-Pure kit (Omega). Differing from previous work, a modified constant region was used. As in previous work, the bases at the first two positions in the gRNA were always substituted for guanine to accommodate preferences of T7 RNA polymerase regardless of the genomic sequence. All gRNAs were transcribed with the Megascript T7 kit (Ambion), DNase treated, and purified by ammonium acetate/ethanol precipitation. RNA concentration was determined using Nanodrop spectrophotometer and agarose gel electrophoresis. Activity of gRNAs was tested by T7 Endonuclease I assays on the PCR products encompassing the gRNA target site, amplified from HotSHOT-extracted genomic DNA (Meeker et al., 2007) from pools of 20 embryos grown to 24-30 hpf following microinjection with gRNA and catalytically active Cas9.

Fish husbandry and microinjection

Zebrafish TLAB embryos were dechorionated with pronase at one-cell stage. The embryos were injected with 300 pg of mRNA for each dCas9 construct and an approximately equimolar mix of the gRNAs being tested in an excess (Gagnon et al.,

2014; Hwang et al., 2013) amount (300-500 pg). For expression studies, embryos were injected at 1-cell stage with different amounts of Squint mRNA: 0.1, 0.4, 1.6 or 4pg.

RT-qPCR

For gene expression analysis, total RNA was isolated from approximately 20 shield stage zebrafish embryos using QIAgene RNeasy kit. cDNA was made from total RNA using BioRad iScript kit, and analyzed using SYBR[®] Green I (BioRad) and DNA Engine Opticon System (MJ Research). A housekeeping gene β -actin was used for normalization across samples.

In situ hybridization

In situ hybridization on whole mount embryos was carried out using standard methods. Embryos were fixed in 4% formaldehyde overnight, and then dehydrated, rehydrated, and stained. Following staining, embryos were cleared in benzyl benzoate/benzyl alcohol (2:1 vol/vol) and imaged.

REFERENCES

- Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S.L., Lander, E.S., 2006. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* 125, 315–326. doi:10.1016/j.cell.2006.02.041
- Bertero, A., Madrigal, P., Galli, A., Hubner, N.C., Moreno, I., Burks, D., Brown, S., Pedersen, R.A., Gaffney, D., Mendjan, S., Pauklin, S., Vallier, L., 2015. Activin/nodal signaling and NANOG orchestrate human embryonic stem cell fate decisions by controlling the H3K4me3 chromatin mark. *Genes Dev.* 29, 702–717. doi:10.1101/gad.255984.114
- DaCosta Byfield, S., Major, C., Laping, N.J., Roberts, A.B., 2004. SB-505124 is a selective inhibitor of transforming growth factor-beta type I receptors ALK4, ALK5, and ALK7. *Mol. Pharmacol.* 65, 744–752. doi:10.1124/mol.65.3.744
- Dahle, Ø., Kumar, A., Kuehn, M.R., 2010. Nodal signaling recruits the histone demethylase Jmjd3 to counteract polycomb-mediated repression at target genes. *Sci. Signal.* 3, ra48. doi:10.1126/scisignal.2000841
- De Groote, M.L., Verschure, P.J., Rots, M.G., 2012. Epigenetic Editing: Targeted rewriting of epigenetic marks to modulate expression of selected target genes. *Nucleic Acids Res.* 40, 10596–10613. doi:10.1093/nar/gks863
- Dubrulle, J., Jordan, B.M., Akhmetova, L., Farrell, J.A., Kim, S.-H., Solnica-Krezel, L., Schier, A.F., 2015. Response to Nodal morphogen gradient is determined by the kinetics of target gene induction. *Elife* 4. doi:10.7554/eLife.05042
- Feldman, B., Gates, M.A., Egan, E.S., Dougan, S.T., Rennebeck, G., Sirotkin, H.I., Schier, A.F., Talbot, W.S., 1998. Zebrafish organizer development and germ-layer formation require nodal-related signals. *Nature* 395, 181–185. doi:10.1038/26013
- Gagnon, J.A., Valen, E., Thyme, S.B., Huang, P., Ahkmetova, L., Pauli, A., Montague, T.G., Zimmerman, S., Richter, C., Schier, A.F., 2014. Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS One* 9. doi:10.1371/journal.pone.0098186
- Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. doi:10.1038/nmeth.1318
- Hilton, I.B., Ippolito, A.M.D., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., Gersbach, C. a, 2015. Epigenome editing by a CRISPR-Cas9-based

- acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* 33. doi:10.1038/nbt.3199
- Hong, S., Cho, Y.-W., Yu, L.-R., Yu, H., Veenstra, T.D., Ge, K., 2007. Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases. *Proc. Natl. Acad. Sci. U. S. A.* 104, 18439–18444. doi:10.1073/pnas.0707292104
- Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.-R.J., Joung, J.K., 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 227–9. doi:10.1038/nbt.2501
- Kearns, N.A., Pham, H., Tabak, B., Genga, R.M., Silverstein, N.J., Garber, M., Maehr, R., 2015. Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Meth* 12, 401–403.
- Keung, A.J., Bashor, C.J., Kiriakov, S., Collins, J.J., Khalil, A.S., 2014. Using targeted chromatin regulators to engineer combinatorial and spatial transcriptional regulation. *Cell* 158, 110–20. doi:10.1016/j.cell.2014.04.047
- Kim, S.W., Yoon, S.J., Chuong, E., Oyolu, C., Wills, A.E., Gupta, R., Baker, J., 2011. Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. *Dev. Biol.* 357, 492–504.
- Meeker, N.D., Hutchinson, S.A., Ho, L., Trede, N.S., 2007. Method for isolation of PCR-ready genomic DNA from zebrafish tissues. *Biotechniques* 43, 610–614. doi:10.2144/000112619
- Mendenhall, E.M., Williamson, K.E., Reyon, D., Zou, J.Y., Ram, O., Joung, J.K., Bernstein, B.E., 2013. Locus-specific editing of histone modifications at endogenous enhancers. *Nat. Biotechnol.* 31, 1133–6. doi:10.1038/nbt.2701
- Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., Nusbaum, C., Lander, E.S., Bernstein, B.E., 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553–560. doi:10.1038/nature06008
- Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G.A., Stewart, R., Thomson, J.A., 2007. Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* 1, 299–312. doi:10.1016/j.stem.2007.08.003
- Pédélecq, J.-D., Cabantous, S., Tran, T., Terwilliger, T.C., Waldo, G.S., 2006. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* 24, 79–88. doi:10.1038/nbt1172

- Rush, M., Appanah, R., Lee, S., Lam, L.L., Goyal, P., Lorincz, M.C., 2009. Targeting of EZH2 to a defined genomic site is sufficient for recruitment of Dnmt3a but not de novo DNA methylation. *Epigenetics* 4, 404–414. doi:10.4161/epi.4.6.9392
- Schier, A.F., 2007. The maternal-zygotic transition: death and birth of RNAs. *Science* 316, 406–7. doi:10.1126/science.1140693
- Vastenhouw, N.L., Zhang, Y., Woods, I.G., Imam, F., Regev, A., Liu, X.S., Rinn, J., Schier, A.F., 2010. Chromatin signature of embryonic pluripotency is established during genome activation. *Nature* 464, 922–926. doi:10.1038/nature08866
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., Liu, X.S., 2008a. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137. doi:10.1186/gb-2008-9-9-r137
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., Liu, X.S., 2008b. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9, R137. doi:10.1186/gb-2008-9-9-r137

CHAPTER V

Perspectives and outstanding questions

Differential gene expression regulation in response to exposure to different levels of Nodal

One of the most interesting aspects of Nodal signaling is its nature as a morphogen. It is the first identified endogenous vertebrate morphogen, and a lot of research has been dedicated to understanding how the gradient is formed and interpreted (Chen and Schier, 2001; Rogers and Schier, 2011; Schier, 2009). The question of how, exactly, high levels of and/or long exposure to Nodal signals can activate one group of genes, and low levels/short exposure – another, still remains unanswered. A similar question about the mechanism of how Nodal signaling switches function exists in the field of human stem cell biology, where this signal can both maintain pluripotency and induce differentiation. It has been suggested that this is achieved through the existence of ‘synexpression groups’, where Smads regulate different groups of target genes in different manners via specific cofactors. It has, indeed been observed that different cofactors in stem cells show specificity for either pluripotency or differentiation genes (Pauklin and Vallier, 2015).

Whether the different groups of genes that Nodal activates during zebrafish embryogenesis are regulated in this way remains to be seen. Notably, while some Smad2-interacting transcription factors are maternally provided (FoxH1, EomesA, Oct1), others are expressed later in the development, often in response to Nodal signaling (Mixer, Mezzo). The fact that different groups of transcription factors appear at different stages supports the idea of synexpression groups in zebrafish, especially for groups of genes requiring different lengths of exposure to Nodal.

Proteome-wide analyses are yet to reveal the full list of Smad2 binding partners, and we are likely to be underestimating the number of factors that interact with Smad2 directly or indirectly. Smad2 co-precipitation analyses at different developmental stages and in embryos treated with different levels of Nodal will be needed to clarify these questions.

The role of repressive transcription factors in the regulation of Nodal target genes has also been possibly underestimated. Negative feedback regulation can play important roles in establishing embryonic patterns. For example, I found that the binding sites for Zic2/3 and Osr1 transcription factors are very abundant in FoxH1 and Smad2 binding elements. While it has been observed that Zic2/3 and Osr1 proteins play a role in dorso-ventral and mesendodermal development in fish and other organisms, in depth analysis of how these factors interact with Smad2 has not been done (Cast et al., 2012; Grinblat and Sive, 2001; Mudumana et al., 2008; Terashima et al., 2014; Winata et al., 2013).

The molecular function of Smad2 itself is unclear. We do not know whether the only role of Smad2 is to assemble the transcriptional complexes required for gene activation, or whether it directly controls the activation of epigenetic modifiers and transcriptional apparatus.

Function and action mechanism of FoxH1 as a pioneer factor

The fact that FoxH1 binds to enhancers without Nodal signaling and later recruits Smad2 to these loci suggests that FoxH1 plays the role of a pioneer factor. There are, however, still tests to be done to confirm that FoxH1 can, indeed, bind to DNA independently of other factors. Many, though not all, FoxH1 binding sites possess an Oct-family consensus binding sequence. It is possible, therefore, that

FoxH1 can only bind to DNA at the positions where Oct4, for example, is already bound. To test this possibility, we would need to do FoxH1 pull-down experiments in Oct4 mutant background. If FoxH1 is, indeed, a pioneer factor, it should be able to bind to some if not all of its binding loci even in these embryos.

Since previously identified pioneer factors often play roles in modifying/remodeling histones and changing/maintaining DNA methylation status, we should also be able to observe chromatin mark changes associated with FoxH1 binding. ChIP-Seq experiments in FoxH1 mutant background, as well as co-immunoprecipitation experiments to identify potential chromatin modifiers that interact with FoxH1, should be able to answer some of these questions,.

Developmental studies have identified many transcription factors that are likely to function as pioneer factors. One such transcription factor MyoD can initiate the myogenic program when introduced into many types of terminally differentiated cells (Weintraub et al., 1989). It has since been shown that many such pioneer factors can initiate reprogramming of cells (Iwafuchi-Doi and Zaret, 2014). There is significant interest in using these direct reprogramming approaches to create cells of biomedical interest, including mesendoderm-derived tissues, such as liver hepatocytes and pancreatic b cells. If FoxH1 does prove to be capable of affecting cell reprogramming, possibly in the presence of Nodal signals, this would be an important finding for the biomedical field.

The interaction of Nodal signaling and chromatin

Chapter IV of this work addressed the interaction between Nodal signaling and H3K4me3/K27me3 marks on its target genes. While we shed some light on what happens to these marks in the presence or absence of this signal, there are

still a lot of answered questions. We still do not understand the sequence of events that happen at the chromatin level. A number of different scenarios are possible here, one where upon zygotic genome activation most of the Nodal target genes acquire a bivalent mark, which then resolves either into a silencing or activating one, depending on Nodal exposure. It is also possible that many of these genes possess monovalent H3K4me3 mark, and then either acquire the silencing modification or not, if Nodal is detected. The third possibility is that these genes are not marked at first, and then acquire H3K4me3 or H3K27me3, depending on the signaling status. Very few genes have been shown to possess the monovalent H3K27me3 mark, so it not likely that such is the starting status of the target genes (Vastenhouw et al., 2010).

While many of the Nodal regulated genes possess both H3K4me3 and H3K27me3 mark in wild-type embryos right after zygotic activation, it is not clear whether those are truly bivalent marks, or the result of heterogeneity of the cell population (Vastenhouw et al., 2010). Sequential ChIP experiments would be required to come to a definite conclusion on this matter.

A large number of chromatin modifications are involved in gene regulation, and are certain to also be involved in the Nodal target gene activation and repression (Ross and Hill, 2008). We will need to delineate the sequence of events at the chromatin level to really understand the roles of different transcription factors in expression regulation. Cas9-based tools should be able to help with a lot of these studies, as exemplified by our work.

Biomedical applications: detangling the dual function of Nodal signaling in stem cells

Nodal signaling controls various mechanisms in different model organisms and in a diversity of cell types. In embryonic stem cells it is required for self-renewal as well as mesendoderm development. It is difficult to untangle these two roles of Nodal in stem cell culture studies, as gain or loss of this signal can lead to uncontrolled proliferation or quiescence. On the other hand, studying the role of Nodal signaling in the early zebrafish embryo allows us to focus on its role in differentiation. Since many cancers are associated with defects in Nodal regulation of self-renewal and cell-cycle control, a more complete picture of the mechanistic aspects of Nodal signaling in zebrafish could help us separate these functions and develop novel therapeutic targets for the treatment of cancer.

Until a few years ago there was a significant limitation to the types of experiments that could be done in zebrafish because many tools for genomic manipulation don't work very well in this system. Targeted mutations for example were almost impossible to do. With the discovery of Cas9-based genomic manipulations, however, the field changed completely. A new mutant for a specific gene can be created in just a few months. Cas9 can also be used for targeted recruitment of various factors to DNA, including histone-modifying factors, as we demonstrated in chapter IV. Zebrafish embryos are also particularly well suited for large-scale screening of therapeutic drugs. Taken together, these facts indicate that studying Nodal signaling in early zebrafish development will continue to be important from biomedical standpoint as well as for our basic understanding of biology.

REFERENCES

- Cast, A.E., Gao, C., Amack, J.D., Ware, S.M., 2012. An essential and highly conserved role for *Zic3* in left-right patterning, gastrulation and convergent extension morphogenesis. *Dev. Biol.* 364, 22–31. doi:10.1016/j.ydbio.2012.01.011
- Chen, Y., Schier, a F., 2001. The zebrafish Nodal signal *Squint* functions as a morphogen. *Nature* 411, 607–610. doi:10.1038/35079121
- Grinblat, Y., Sive, H., 2001. *zic* gene expression marks anteroposterior pattern in the presumptive neur ectoderm of the zebrafish gastrula. *Dev. Dyn.* 222, 688–693. doi:10.1002/dvdy.1221
- Iwafuchi-Doi, M., Zaret, K.S., 2014. Pioneer transcription factors in cell reprogramming. *Genes Dev.* doi:10.1101/gad.253443.114
- Mudumana, S.P., Hentschel, D., Liu, Y., Vasilyev, A., Drummond, I. a, 2008. Odd Skipped Related1 Reveals a Novel Role for Endoderm in Regulating Kidney Versus Vascular Cell Fate. *Development* 135, 3355–67. doi:10.1242/dev.022830
- Pauklin, S., Vallier, L., 2015. Activin/Nodal signalling in stem cells. *Development* 142, 607–19. doi:10.1242/dev.091769
- Rogers, K.W., Schier, A.F., 2011. Morphogen Gradients: From Generation to Interpretation. *Annu. Rev. Cell Dev. Biol.* 27, 377–407. doi:10.1146/annurev-cellbio-092910-154148
- Ross, S., Hill, C.S., 2008. How the Smads regulate transcription. *Int. J. Biochem. Cell Biol.* doi:10.1016/j.biocel.2007.09.006
- Schier, A.F., 2009. Nodal morphogens. *Cold Spring Harb. Perspect. Biol.* doi:10.1101/cshperspect.a003459
- Terashima, A. V., Mudumana, S.P., Drummond, I.A., 2014. Odd skipped related 1 is a negative feedback regulator of nodal-induced endoderm development. *Dev. Dyn.* 243, 1571–1580. doi:10.1002/dvdy.24191
- Vastenhouw, N.L., Zhang, Y., Woods, I.G., Imam, F., Regev, A., Liu, X.S., Rinn, J., Schier, A.F., 2010. Chromatin signature of embryonic pluripotency is established during genome activation. *Nature* 464, 922–926. doi:10.1038/nature08866
- Weintraub, H., Tapscott, S.J., Davis, R.L., Thayer, M.J., Adam, M.A., Lassar, A.B., Miller, A.D., 1989. Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of *MyoD*. *Proc. Natl. Acad. Sci. U. S. A.* 86, 5434–8. doi:10.1073/pnas.86.14.5434

Winata, C.L., Kondrychyn, I., Kumar, V., Srinivasan, K.G., Orlov, Y., Ravishankar, A., Prabhakar, S., Stanton, L.W., Korzh, V., Mathavan, S., 2013. Genome Wide Analysis Reveals Zic3 Interaction with Distal Regulatory Elements of Stage Specific Developmental Genes in Zebrafish. PLoS Genet. 9.
doi:10.1371/journal.pgen.1003852